

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 07:48:47 ; Search time 1.76303 Seconds
(without alignments)
604.643 Million cell updates/sec

Title: US-09-857-308-5
Perfect score: 37
Sequence: 1 LYQAVATI 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	100.0	8	21 AAB03883	Human ART-1 peptid
2	37	100.0	8	22 AAB46941	Human ART-3 deri
3	37	100.0	9	21 AAB03886	Human ART-1 peptid
4	37	100.0	363	22 AAG98646	Human cell death p
5	37	100.0	403	21 AAB45216	Gene 41 human secr
6	37	100.0	404	21 AAB45217	Human secreted pro
7	37	100.0	412	22 AAG98644	Human cell death p
8	37	100.0	414	21 AAB03880	Human ART-1 protei
9	37	100.0	419	22 AAM25845	Human protein sequ
10	32	86.5	100	23 ABB28637	Streptococcus poly

11	32	86.5	273	21	AA909399	Cenarchaeum symbio
12	31	83.8	269	20	AA909233	Cenarchaeum symbio
13	30	81.1	9	20	AA947894	Immunogenic peptid
14	30	81.1	9	22	AA22249	HIV peptide SEQ ID
15	30	81.1	9	22	AAU12541	Human HIV-1 Th-CTL
16	30	81.1	20	19	AAW55857	HIV p17 gag protei
17	30	81.1	20	22	AAU12335	Peptide fragment o
18	30	81.1	24	22	AAU12487	HIV Th-CTL peptide
19	30	81.1	24	23	AAU70250	HIV Th-CTL P17 epi
20	30	81.1	123	23	AAU10483	HIV Th-CTL epitope
21	30	81.1	139	22	ABB10679	Human pancreatic c
22	30	81.1	139	22	AA92721	Human digestive sy
23	30	81.1	247	22	AA92296	Human protein sequ
24	30	81.1	312	18	AAW20350	H. pylori outer me
25	30	81.1	312	18	AAW20857	H. pylori outer me
26	30	81.1	449	22	AA92706	Human protein sequ
27	30	81.1	577	22	AA93792	Human polypeptide,
28	30	81.1	756	22	AA087329	Novel central nerv
29	30	81.1	814	22	AA43565	Human polypeptide
30	29	78.4	9	20	AA47900	Immunogenic peptid
31	29	78.4	95	22	AAU14913	Novel bone marrow
32	29	78.4	158	22	AAU42587	Propionibacterium
33	29	78.4	269	17	AA88725	LIF3. Mycoplasma
34	29	78.4	560	20	AA99104	Human EAAT5 glutam
35	29	78.4	560	21	AA52196	Excitatory amino a
36	29	78.4	560	22	AAE10884	Human excitatory a
37	29	78.4	560	22	AAE11030	Human excitatory a
38	29	78.4	564	20	AA99105	Protein sequence S
39	29	78.4	564	21	AA52197	Excitatory amino a
40	29	78.4	564	22	AAE10885	Human excitatory a
41	29	78.4	572	17	AAW02256	Mouse L-glutamate
42	29	78.4	574	18	AAW26598	Human glutamate tr
43	29	78.4	574	19	AAW58553	Human excitatory a
44	29	78.4	574	20	AA42551	Human astrocyte wi
45	29	78.4	574	20	AA28296	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAB03883
ID AAB03883 standard; protein; 8 AA.
XX
AC AAB03883;
XX
DT 26-OCT-2000 (first entry)
XX
DE Human ART-1 peptide fragment SEQ ID #5.
XX
DE Human; tumour antigen protein; ART-1; HLA antigen; cytostatic;
KW bound cytotoxic T cell; tumour; cancer; remedy; prevent; diagnose.
XX
OS Homo sapiens.
PN WO200032770-A1.
XX
PD 08-JUN-2000.
XX
PF 30-NOV-1999; 99WO-JP06682.
XX
PR 01-DEC-1998; 98JP-0341253.
XX
PA (SUMU) SUMITOMO PHARM CO LTD.
XX
PI (ITOH/) ITOH K.
XX
PI Itoh K, Gomi S;
DR WPI; 2000-412318/35.
XX
PT Novel tumor antigen protein ART-1, tumor antigen peptide originating
PT from it, their derivatives, and DNAs, applicable in vivo or in vitro as
PT remedies, preventives and diagnostics for tumors -

XX Claim 11; Page 35; 59pp; Japanese.

XX The invention relates to a novel human tumour antigen protein, ART-1.

CC Included in the invention are polynucleotide sequences encoding the ART-1

CC protein, and mutated ART-1 proteins which when broken down

CC intracellularly produce a tumour antigen peptide that can recognise HLA

CC antigen and bound cytotoxic T cells. Antibodies which specifically

CC recognise ART-1 and its derivative peptides, are also included in the

CC invention. ART-1 exhibits cytostatic activity. The tumour antigen

CC protein, tumour antigen peptide originating from it, their derivatives,

CC and DNAs are applicable in vivo or in vitro as remedies, preventives and

CC diagnostics for tumours.

CC The present sequence represents a fragment of the ART-1 protein, used in

CC the invention.

XX Sequence 8 AA;

XX Query Match 100.0%; Score 37; DB 21; Length 8;

XX Best Local Similarity 100.0%; Pred. No. 7.8e+05;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

Db 1 LYQAVATI 8

RESULT 2

AAB46941

ID AAB46941 standard; peptide; 8 AA.

XX AAB46941;

XX 04-MAY-2001 (first entry)

XX Human ART-3 derived tumor antigenic peptide SEQ ID 27.

XX Tumor antigenic peptide; Interferon; IFN; antigen-specific T cell;

KW virucide; cytostatic; gene therapy; IFN-alpha; cytotoxic T cell;

KW viral infectious disease; ART-1; human.

XX Homo sapiens.

XX EP1074267-A1.

XX 07-FEB-2001.

XX 24-JUL-2000; 2000EP-0306263.

XX 22-JUL-1999; 99JP-0207687.

XX (SUMU) SUMITOMO PHARM CO LTD.

XX Takasu H, Gotoh M, Yamaoka T;

XX WPI; 2001-193144/20.

XX Use of antigenic proteins, peptides, interferon or their encoding DNA,

PT in the manufacture of an agent for the induction of antigen-specific T

PT cells

XX Disclosure; Page 18; 25pp; English.

XX This invention describes the novel use of interferons (IFNs) or DNAs

CC capable of expressing the interferons and/or antigenic proteins (AP),

CC antigenic peptides derived from the proteins or DNAs capable of

CC expressing the antigenic proteins or peptides, in the manufacture of

CC an agent for induction of antigen-specific T cells. The products of

CC the invention have virucide and cytostatic activity and can be used for

CC gene therapy or as inducers of antigen-specific T cells. The action of

CC interferon-alpha (IFN-alpha) in a system for inducing specific cytotoxic

CC T cell (CTL) by administering an antigenic peptide in an incomplete

CC Freund's adjuvant (IFA) emulsion preparation form was evaluated. IFNs (or

CC DNA encoding IFNs) are useful in the manufacture of a medicament for

CC inducing antigen-specific T cells in an individual who has been

CC administered with AP (or DNA encoding AP) or vice versa. The medicament

CC is useful for the treatment or prophylaxis of a tumor or a viral

CC infectious disease.

XX Sequence 8 AA;

XX Query Match 100.0%; Score 37; DB 22; Length 8;

XX Best Local Similarity 100.0%; Pred. No. 7.8e+05;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

Db 1 LYQAVATI 8

RESULT 3

AAB03886

ID AAB03886 standard; protein; 9 AA.

XX AAB03886;

XX 26-OCT-2000 (first entry)

XX Human ART-1 peptide fragment SEQ ID #8.

XX Human; tumour antigen protein; ART-1; HLA antigen; cytostatic;

KW bound cytotoxic T cell; tumour; cancer; remedy; prevent; diagnose.

XX Homo sapiens.

XX WO200032770-A1.

XX 08-JUN-2000.

XX 30-NOV-1999; 99WO-JP06682.

XX 01-DEC-1998; 98JP-0341253.

XX (SUMU) SUMITOMO PHARM CO LTD.

XX (ITOHI) ITOH K.

XX Itoh K, Gomi S;

XX WPI; 2000-412318/35.

XX Novel tumor antigen protein ART-1, tumor antigen peptide originating

PT from it, their derivatives, and DNAs, applicable in vivo or in vitro as

PT remedies, preventives and diagnostics for tumors

XX Claim 11; Page 51; 59pp; Japanese.

XX The invention relates to a novel human tumour antigen protein, ART-1.

CC Included in the invention are polynucleotide sequences encoding the ART-1

CC protein, and mutated ART-1 proteins which when broken down

CC intracellularly produce a tumour antigen peptide that can recognise HLA

CC antigen and bound cytotoxic T cells. Antibodies which specifically

CC recognise ART-1 and its derivative peptides, are also included in the

CC invention. ART-1 exhibits cytostatic activity. The tumour antigen

CC protein, tumour antigen peptide originating from it, their derivatives,

CC and DNAs are applicable in vivo or in vitro as remedies, preventives and

CC diagnostics for tumours.

CC The present sequence represents a fragment of the ART-1 protein, used in

CC the invention.

XX Sequence 9 AA;

XX Query Match 100.0%; Score 37; DB 21; Length 9;

XX Best Local Similarity 100.0%; Pred. No. 7.8e+05;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

DB 1 LYQAVATI 8

RESULT 4

AAAG98646

ID AAC98646 standard; Protein: 363 AA.

XX AAC98646;

AC AAC98646;

XX 21-SEP-2001 (first entry)

DT Human cell death protective cDNA clone CNI-00714 ORF3 protein, SEQ.81.

DE

XX Cell death protective; apoptosis; necrosis; human; drug screening;

XX cell death-associated disorder; central nervous system disorder;

KW psychiatric disorder; neurological disorder; ischaemia-related disorder;

KW stroke; cerebral infarction; ischaemic encephalopathy;

KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;

KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;

KW vascular disease; ophthalmological disorder; diabetic retinopathy;

KW macular degeneration; hypertension; myocardial infarction;

KW atherosclerosis; respiratory disorder; asthma; transgenic animal;

KW chronic obstructive pulmonary disease; neoplastic condition; cancer;

KW benign tumour; anaemia; gastrointestinal disorder; gastritis;

KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;

KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;

KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;

KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.

XX Homo sapiens.

OS

XX WO200145638-A2.

PN

XX 28-JUN-2001.

PD

XX 11-DEC-2000; 2000WO-US33547.

XX

XX 14-DEC-1999; 99US-0461697.

PR

XX (COGE-) COGENT NEUROSCIENCE INC.

XX

PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;

XX WPI; 2001-390297/41.

DR N-PSDB; AAH84170, AAH84173.

DR

XX Novel protective sequence polynucleotides and polypeptides, used to

PT identify modulators of their expression and activity, which are used in

PT to treat central nervous system conditions, diseases and disorders -

XX

XX Claim 1; Fig 6C; 325pp; English.

PS

XX Sequences AAH84132-AAH84370 represent human nucleic acid sequences which

CC protect against cell death (i.e., apoptosis or necrosis). Sequences

CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,

CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,

CC while the remaining nucleic acid sequences within the range given above

CC represent the open reading frames (ORFs) of these cDNA clones. Sequences

CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death

CC protective ORFs. The cell death protective cDNA clones are able to

CC prevent, delay or reverse progression through the apoptotic or necrotic

CC pathways when injected into a cell predisposed to or undergoing cell

CC death. The cell death protective nucleic acids and polypeptides can be

CC used in the diagnosis and treatment of disorders associated with cell

CC death, and to screen for compounds which modulate their activity or

CC expression. Such modulators, preferably a small organic molecule, an

CC antibody, a ribozyme, or an antisense molecule, can also be used to treat

CC cell death-related diseases. Such diseases include those associated with

CC the central nervous system including psychiatric or neurological

CC disorders, especially ischaemia-related conditions such as strokes, and

CC also includes neurodegenerative disorders such as Alzheimer's disease,

CC Huntington's disease, or Parkinson's disease. The modulators may also be

used to treat infections such as meningitis, malaria, or trypanosomiasis;

vascular diseases such as ischaemic encephalopathy or cerebral

infarction; eye conditions such as diabetic retinopathy or macular

degeneration; hypertension; myocardial infarction; atherosclerosis;

respiratory conditions such as asthma or chronic obstructive pulmonary

disease; neoplastic conditions such as cancers or benign tumours; blood

cell conditions such as anaemia; gastrointestinal conditions such as

gastritis or ulcerative colitis; liver conditions such as biliary

cirrhosis; kidney disorders such as glomerulonephritis; cystitis;

endometriosis; endocrine disorders such as Grave's disease or Hashimoto's

thyroiditis; skin conditions such as dermatitis or urticaria; or immune

system disorders such as acquired immunodeficiency syndrome (AIDS). The

nucleic acids may additionally be used to generate animal models of

cell death-associated disorders. The present sequence represents a

cell death protective polypeptide.

XX

XX Sequence 363 AA;

Query Match 100.0%; Score 37; DB 22; Length 363;

Best Local Similarity 100.0%; Pred. No. 2.6;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8

DB 107 LYQAVATI 114

RESULT 5

AAAB45216

ID AAB45216 standard; Protein: 403 AA.

XX AAB45216;

AC

XX 12-FEB-2001 (first entry)

DT

XX Gene 41 human secreted protein homologous amino acid sequence #157.

DE

XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;

KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;

KW neurotropic; neuroprotective; antibacterial; virucide; fungicide; cancer;

KW ophthalmological; autoimmune disease; hyperproliferative disorder;

KW cardiovascular disorder; cerebrovascular disorder; wound healing;

KW nervous system disorder; aging; chemotaxis.

XX Homo sapiens.

OS

XX WO200058467-A1.

PN

XX 05-OCT-2000.

PD

XX 22-MAR-2000; 2000WO-US07505.

PF

XX 26-MAR-1999; 99US-0126502.

PR

XX 17-DEC-1999; 99US-0172410.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-611712/58.

DR

XX Nucleic acid molecules encoding human secreted proteins, used in

PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and

PT Parkinson's diseases and cancers -

XX

XX Disclosure; Page 60-61; 440pp; English.

PS

XX Polynucleotide sequences AAC80531-C80580 represent cDNA encoding human

CC secreted proteins AAB45120-B45169. Sequences AAB45170-B45225 represent

CC alternative polypeptides encoded by the genes, and amino acid sequences

CC to which they are homologous. The genes and proteins have activities

CC dependent on the tissues and cells in which they are expressed. Examples

CC of their activities include immunosuppressive; antiarthritic;

CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 CC cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
 CC fungicide; and ophthalmological. The secreted proteins, polynucleotides,
 CC antagonists and agonists may be useful in treating, preventing and/or
 CC diagnosing diseases and disorders such as autoimmune diseases
 CC e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms
 CC of the breast or liver, cardiovascular disorders e.g. cardiac arrest,
 CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous
 CC system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
 CC The polypeptides can also be used to aid wound healing and epithelial
 CC cell proliferation, to prevent skin aging due to sunburn, to maintain
 CC organs before transplantation, for supporting cell culture of primary
 CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
 CC also be used as a food additive or preservative to increase or decrease
 CC storage capabilities. AAC80522-C80530 and AAB45119 represent sequences
 CC used in the isolation and characterisation of the genes and proteins of
 CC the invention.
 XX
 SQ Sequence 403 AA;
 Query Match 100.0%; Score 37; DB 21; Length 403;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LYQAVATI 8
 Db 154 LYQAVATI 161
 |||||
 RESULT 6
 AAB45217
 ID AAB45217 standard; Protein; 404 AA.
 XX AAB45217;
 AC AAB45217;
 DT 12-FEB-2001 (first entry)
 DE Human secreted protein sequence encoded by gene 41 SEQ ID NO:158.
 DE
 XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
 KW neurotropic; neuroprotective; antibacterial; virucide; fungicide; cancer;
 KW ophthalmological; autoimmune disease; hyperproliferative disorder;
 KW cardiovascular disorder; cerebrovascular disorder; wound healing;
 KW nervous system disorder; aging; chemotaxis.
 XX
 OS Homo sapiens.
 XX
 XX WO2000058467-A1.
 PN
 XX
 PD 05-OCT-2000.
 XX
 XX 22-MAR-2000; 2000WO-US07505.
 PF
 XX 26-MAR-1999; 99US-0126502.
 PR
 XX 17-DEC-1999; 99US-0172410.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Rosen CA, Ruben SM, Komatsoulis G;
 PI
 XX WPI; 2000-611712/58.
 DR
 XX N-PSDB; AAC80571.
 DR
 XX Nucleic acid molecules encoding human secreted proteins, used in
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 PT Parkinson's diseases and cancers -
 XX
 PS Disclosure; Page 60-61; 440pp; English.
 PS
 XX Polynucleotide sequences AAC80531-C80580 represent cDNA encoding human
 CC secreted proteins AAB45120-B45169. Sequences AAB45170-B45225 represent

CC alternative polypeptides encoded by the genes, and amino acid sequences
 CC to which they are homologous. The genes and proteins have activities
 CC dependent on the tissues and cells in which they are expressed. Examples
 CC of their activities include immunosuppressive; antiarthritic;
 CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 CC cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
 CC fungicide; and ophthalmological. The secreted proteins, polynucleotides,
 CC antagonists and agonists may be useful in treating, preventing and/or
 CC diagnosing diseases and disorders such as autoimmune diseases
 CC e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms
 CC of the breast or liver, cardiovascular disorders e.g. cardiac arrest,
 CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous
 CC system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
 CC The polypeptides can also be used to aid wound healing and epithelial
 CC cell proliferation, to prevent skin aging due to sunburn, to maintain
 CC organs before transplantation, for supporting cell culture of primary
 CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
 CC also be used as a food additive or preservative to increase or decrease
 CC storage capabilities. AAC80522-C80530 and AAB45119 represent sequences
 CC used in the isolation and characterisation of the genes and proteins of
 CC the invention.
 XX
 SQ Sequence 404 AA;
 Query Match 100.0%; Score 37; DB 21; Length 404;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LYQAVATI 8
 Db 154 LYQAVATI 161
 |||||
 RESULT 7
 AAG98644
 ID AAG98644 standard; Protein; 412 AA.
 XX AAG98644;
 AC AAG98644;
 DT 21-SEP-2001 (first entry)
 DE Human cell death protective cDNA clone CNI-00714 ORF1 protein, SEQ:77.
 DE
 XX Cell death protective; apoptosis; necrosis; human; drug screening;
 KW cell death-associated disorder; central nervous system disorder;
 KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
 KW stroke; cerebral infarction; ischaemic encephalopathy;
 KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
 KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
 KW vascular disease; ophthalmological disorder; diabetic retinopathy;
 KW macular degeneration; hypertension; myocardial infarction;
 KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
 KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
 KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
 KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
 KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
 KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
 KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.
 XX
 OS Homo sapiens.
 XX
 XX WO200145638-A2.
 PN
 XX 28-JUN-2001.
 PD
 XX 11-DEC-2000; 2000WO-US33547.
 PF
 XX 14-DEC-1999; 99US-0461697.
 PR
 XX (COGE-) COGENT NEUROSCIENCE INC.
 PA
 XX Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
 PI

XX WPI: 2001-390297/41.
 DR N-PSDB; AAH84170, AAH84171.
 XX
 PT Novel protective sequence polynucleotides and polypeptides, used to
 PT identify modulators of their expression and activity, which are used in
 PT to treat central nervous system conditions, diseases and disorders -
 XX
 PS Claim 1; Fig 6A; 325pp; English.
 XX
 CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
 CC protect against cell death (i.e., apoptosis or necrosis). Sequences
 CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
 CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
 CC while the remaining nucleic acid sequences within the range given above
 CC represent the open reading frames (ORFs) of these cDNA clones. Sequences
 CC AG98610-AG98829 represent the polypeptides encoded by the cell death
 CC protective ORFs. The cell death protective cDNA clones are able to
 CC prevent, delay or reverse progression through the apoptotic or necrotic
 CC pathways when injected into a cell predisposed to or undergoing cell
 CC death. The cell death protective nucleic acids and polypeptides can be
 CC used in the diagnosis and treatment of disorders associated with cell
 CC death, and to screen for compounds which modulate their activity or
 CC expression. Such modulators, preferably a small organic molecule, an
 CC antibody, a ribozyme, or an antisense molecule, can also be used to treat
 CC cell death-related diseases. Such diseases include those associated with
 CC the central nervous system including psychiatric or neurological
 CC disorders, especially ischaemia-related conditions such as strokes, and
 CC also includes neurodegenerative disorders such as Alzheimer's disease,
 CC Huntington's disease, or Parkinson's disease. The modulators may also be
 CC used to treat infections such as meningitis, malaria, or trypanosomiasis;
 CC vascular diseases such as ischaemic encephalopathy or cerebral
 CC infarction; eye conditions such as diabetic retinopathy or macular
 CC degeneration; hypertension; myocardial infarction; atherosclerosis;
 CC respiratory conditions such as asthma or chronic obstructive pulmonary
 CC disease; neoplastic conditions such as cancers or benign tumours; blood
 CC cell conditions such as anaemia; gastrointestinal conditions such as
 CC gastritis or ulcerative colitis; liver conditions such as biliary
 CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;
 CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's
 CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune
 CC system disorders such as acquired immunodeficiency syndrome (AIDS). The
 CC nucleic acids may additionally be used to generate animal models of
 CC cell death-associated disorders. The present sequence represents a
 CC cell death protective polypeptide.
 XX
 SQ Sequence 412 AA;

Query Match 100.0%; Score 37; DB 22; Length 412;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
 |||||
 Db 156 LYQAVATI 163

RESULT 8
 AAB03880
 ID AAB03880 standard; protein: 414 AA.
 XX
 XX AAB03880;
 AC
 XX
 XX 26-OCT-2000 (first entry)
 DT
 DE Human ART-1 protein sequence.
 DE
 XX Human; tumour antigen protein; ART-1; HLA antigen; cytostatic;
 XX bound cytotoxic T cell; tumour; cancer; remedy; prevent; diagnose.
 XX
 OS Homo sapiens.
 XX
 PN WO200032770-A1.

XX 08-JUN-2000.
 PD
 XX
 PF 30-NOV-1999; 99WO-JP06682.
 XX
 PR 01-DEC-1998; 98JP-0341253.
 XX
 PA (SUMU) SUMITOMO PHARM CO LTD.
 PA (ITOHI/) ITOH K.
 XX
 XX Itoh K, Gomi S;
 XX
 DR WPI: 2000-412318/35.
 DR N-PSDB; AAA62864.
 XX
 PT Novel tumor antigen protein ART-1, tumor antigen peptide originating
 PT from it, their derivatives, and DNAs, applicable in vivo or in vitro as
 PT remedies, preventives and diagnostics for tumors -
 XX
 PS Claim 1; Page 44-46; 59pp; Japanese.
 XX
 CC The invention relates to a novel human tumour antigen protein, ART-1.
 CC Included in the invention are polynucleotide sequences encoding the ART-1
 CC protein, and mutated ART-1 proteins which when broken down
 CC intracellularly produce a tumour antigen peptide that can recognise HLA
 CC antigen and bound cytotoxic T cells. Antibodies which specifically
 CC recognise ART-1 and its derivative peptides, are also included in the
 CC invention. ART-1 exhibits cytostatic activity. The tumour antigen
 CC protein, tumour antigen peptide originating from it, their derivatives, and
 CC DNAs are applicable in vivo or in vitro as remedies, preventives and
 CC diagnostics for tumours.
 CC The present sequence represents the human ART-1 protein sequence.
 XX
 SQ Sequence 414 AA;

Query Match 100.0%; Score 37; DB 21; Length 414;
 Best Local Similarity 100.0%; Pred. No. 3.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
 |||||
 Db 158 LYQAVATI 165

RESULT 9
 AAM25845
 ID AAM25845 standard; protein: 419 AA.
 XX
 XX AAM25845;
 AC
 XX
 XX 16-OCT-2001 (first entry)
 DT
 DE Human protein sequence SEQ ID NO:1360.
 DE
 XX
 XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnery; antilulcer; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200153455-A2.

```

XX PD 26-JUL-2001.
XX PF
XX PR
XX PR 22-DEC-2000; 2000WO-US35017.
XX PR
XX PR 23-DEC-1999; 99US-0471275.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PA (HYSE-) HYSEQ INC.
XX PI
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-457603/49.
XX DR N-PSDB; AAH99786.
XX PT Isolated human polynucleotides encoding polypeptides, useful for the
XX PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX PS Claim 20; Page 281; 1217pp; English.
XX CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
XX CC AAM25963. The proteins can have activities based on the tissues and
XX CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
XX CC antarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
XX CC central nervous system; viricide; anti-HIV; fungicide; antimutagen;
XX CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
XX CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
XX CC antidiabetic; cytotatic; neuroprotective; antidepressant; nootropic;
XX CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XX CC encoding them can be used in gene therapy, antisense therapy and vaccine
XX CC production. The proteins and polynucleotides are useful for screening for
XX CC agonists or antagonists of a protein and for the treatment and diagnosis
XX CC of disorders associated with the activity of a protein e.g. inflammation,
XX CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XX CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
XX CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
XX CC osteoporosis, severe combined immunodeficiency, eczema, allergic
XX CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
XX CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
XX CC neurological disorders.
XX SQ Sequence 419 AA;
Query Match 100.0%; Score 37; DB 22; Length 419;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LYQAVATI 8
Db 163 LYQAVATI 170
RESULT 10
ABP28637
ID ABP28637 standard; Protein; 100 AA.
XX AC
XX AC ABP28637;
XX DT
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 6450.
XX KW Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus pyogenes.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.

```

```

XX 29-OCT-2001; 2001WO-GB04789.
XX 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN69268.
XX PT New Streptococcus protein for the treatment or prevention of infection
XX PT or disease caused by Streptococcus bacteria, such as meningitis, and
XX PT for detecting a compound that binds to the protein -
XX PS Claim 1; Page 3807; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins.
XX SQ Sequence 100 AA;
Query Match 86.5%; Score 32; DB 23; Length 100;
Best Local Similarity 87.5%; Pred. No. 8.8;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 LYQAVATI 8
Db 76 LYQEVATI 83
RESULT 11
AA90939
ID AA90939 standard; Protein; 273 AA.
XX AC
XX AC AA90939;
XX DT
XX DT 30-AUG-2000 (first entry)
XX DE Cenarchaeum symbiosum open reading frame protein sequence SEQ ID NO:56.
XX KW Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;
XX KW characterisation; archae; therapeutic; industrial; laboratory.
XX OS Cenarchaeum symbiosum.
XX PN WO200018909-A2.
XX PD 06-APR-2000.
XX PF 29-SEP-1999; 99WO-US22752.
XX PR 29-SEP-1998; 98US-0102294.

```

CC	and proteins isolated from the non-thermophilic crenarchaeote
CC	Cenarchaeum symbiosum. The nucleic acids and proteins identified in
CC	the present invention are useful in characterising the physiology of
CC	certain archae and can be used in therapeutic, industrial or laboratory
CC	techniques. AAA55227 to AAA55260 represent promoter sequences from
CC	Cenarchaeum symbiosum. AAA55261 to AAA55269 represent PCR primers and
CC	probes used in examples from the present invention.
XX	
SQ	Sequence 269 AA;
	Query Match 83.8%; Score 31; DB 21; Length 269;
	Best Local Similarity 75.0%; Pred. No. 47;
	Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0
Oy	1 LYQAVATI R
Dd	174 LYQAVPTV IRI
	:
RESULT 13	
ID	AAAY47894 standard; Peptide; 9 AA.
AC	AAAY47894;
XX	
DT	01-DEC-1999 (first entry)
XX	
DE	Immunogenic peptide having a human leukocyte antigen binding motif #2505.
XX	
KW	Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW	immune response; T cell activation; major histocompatibility complex;
KW	cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW	prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW	vaccine; immunisation.
XX	
OS	Synthetic.
OS	Homo sapiens.
XX	
PN	WO9945954-A1.
XX	
PD	16-SEP-1999.
XX	
Pf	13-MAR-1998; 98WO-US05039.
XX	
PR	13-MAR-1998; 98WO-US05039.
XX	
PA	(EPIM-) EPIMUNE INC.
XX	
PI	Sette A, Kubo RT, Sidney J, Cells E, Grey HM, Southwood S;
XX	
DR	WPI; 1999-551214/46.
XX	
PT	New immunogenic peptides with HLA binding motif, useful in treatment
PT	and diagnosis of cancers and viral diseases -
XX	
PS	Claim 1; Page 126; 150pp; English.
XX	
CC	AAAY5390 to AAAY48214 represent specifically claimed immunogenic peptides
CC	having a human major histocompatibility complex (MHC) Class I (also
CC	known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC	peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC	HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC	response against the antigen from which the peptide is derived.
CC	Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC	normally induced by an antigen in the form of a peptide fragment bound
CC	to a HLA molecule, rather than the intact foreign antigen itself, and
CC	are particularly important in tumour rejection and in fighting viral
CC	infections. The peptides are therefore useful therapeutically to treat
CC	or prevent viral infections and cancers in mammals (especially humans)
CC	e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC	They can be administered as vaccines to elicit an immune response in
CC	individuals susceptible or otherwise at risk of viral infection or
CC	cancer, or used to treat chronic or acute conditions. They are also

CC useful diagnostically, and can be used to induce a cytotoxic T cell
 CC response, by contacting a cytotoxic T cell with the peptide e.g. to
 CC produce CTLs ex vivo for infusion back into a patient. The
 CC polynucleotides encoding the immunogenic peptides are also useful
 CC therapeutically and for immunisation as above.

XX Sequence 9 AA;

Query Match 81.1%; Score 30; DB 20; Length 9;
 Best Local Similarity 75.0%; Pred. No. 7.8e+05;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
 |||||
 Db 2 LYNNAVTL 9

RESULT 14

.AAM22249
 ID AAM22249 standard; Peptide; 9 AA.

XX AC AAM22249;

DT 22-OCT-2001 (first entry)

DE HIV peptide SEQ ID NO 134.

XX Cytotoxic T cell lymphocyte; CTL; anti-HIV; viral disease; HIV;
 KW human immunodeficiency virus; vaccine.

XX OS Synthetic.
 OS Human immunodeficiency virus.

XX WO200155177-A2.

XX 02-AUG-2001.

PF 29-JAN-2001; 2001WO-DK00059.

XX 28-JAN-2000; 2000EP-0610017.

PR 31-JAN-2000; 2000US-0179333.

XX (STAT-) STATENS SERUM INST.

XX Fomsgaard A, Brunak S, Buus S, Corbet S, Laemollier SL, Hansen J;
 PI WPI; 2001-476184/51.

XX The generation of cytotoxic T cell lymphocytes epitopes for use in
 PT anti-HIV vaccines

PS Claim 13; Page 53; 383pp; English.

XX The invention relates to identification of cytotoxic T cell lymphocyte
 CC (CTL) epitopes (AAM22116-AAM23484) that generate anti-HIV activity. CTL
 CC are a major protective mechanism against viral diseases. Antibodies may
 CC neutralise extracellular human immunodeficiency virus (HIV) and limit or
 CC prevent infection of cells in the host, but CTL will limit viral
 CC production by killing the cell. The CTL epitopes are useful in medicine,
 CC in the manufacture of vaccines or diagnostic agents.

XX Sequence 9 AA;

Query Match 81.1%; Score 30; DB 22; Length 9;
 Best Local Similarity 75.0%; Pred. No. 7.8e+05;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
 |||||
 Db 2 LYNNAVTL 9

RESULT 15

AAU12541

ID AAU12541 standard; Peptide; 9 AA.

XX AC AAU12541;

DT 27-SEP-2001 (first entry)

DE Human HIV-1 Th-CTL overlapping epitope #48.

XX Human immunodeficiency virus; HIV; immunogen; vaccine; anti-HIV;
 KW human leukocyte antigen; HLA; T-helper epitope; CTL; Th; epitope;
 KW major histocompatibility complex; MHC; cytotoxic T-lymphocyte;
 KW Vaccinia ankara.

XX Homo sapiens.

OS Human immunodeficiency virus type 1.

XX WO200156355-A2.

XX 09-AUG-2001.

PF 05-FEB-2001; 2001WO-US03540.

PR 04-FEB-2000; 2000US-0497497.

XX (UYDU-) UNIV DUKE.

XX Haynes BF, Liao H;

XX WPI; 2001-48827/53.

XX Vaccine for immunizing against human immunodeficiency virus has mixture
 PT or linear array of peptides comprising immunodominant T-helper epitopes
 PT and major histocompatibility complex cytotoxic T-lymphocyte epitopes
 PS Disclosure; Page 27; 33pp; English.

XX The present invention relates to human immunodeficiency virus (HIV)
 CC and in particular to a human leukocyte antigen (HLA)-based HIV vaccine.
 CC The vaccine comprises a mixture or linear array of peptides, or its
 CC variants, where the peptides contain immunodominant T-helper (Th)
 CC epitopes and major histocompatibility complex (MHC) cytotoxic T-
 CC lymphocyte (CTL) epitopes and the linear array of peptides are
 CC preferably expressed in modified Vaccinia ankara. The vaccine is useful
 CC for immunising a patient against HIV and focuses immune response on many
 CC dominant and subdominant CTL epitopes of HIV. DNA or live vectors with
 CC linear arrays of CTL epitopes can be used as either primes or boosts of
 CC peptides or of each other to optimally give CTL anti-HIV responses. The
 CC vaccine induces salutary anti-HIV immune responses. AAU12447-AAU12558
 CC represent the amino acid sequences of the Th-CTL epitopes and HIV
 CC immunogenic peptides used in the invention.

XX Sequence 9 AA;

Query Match 81.1%; Score 30; DB 22; Length 9;
 Best Local Similarity 75.0%; Pred. No. 7.8e+05;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
 |||||
 Db 2 LYNNAVTL 9

Search completed: July 18, 2003, 09:57:08
 Job time : 3.76303 secs

DE 2610524B01RIK protein.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Harai A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Anon H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez L., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK012134; BAB28053.1; -;
DR MGD: MGI:1519445; 2610524B01RIK.
SQ SEQUENCE 412 AA; 45970 MW; B7BF1525BB44F6D5 CRC64;

Query Match 100.0%; Score 37; DB 11; Length 412;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
DB 156 LYQAVATI 163

RESULT 3
O94864 PRELIMINARY; PRT; 414 AA.
AC O94864;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE KIAA0764 protein (Adenocarcinoma antigen ART1).
GN KIAA0764.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20441578; PubMed=10987294;
RA Nishizaka S., Gomi S., Harada K., Oizumi K., Itoh K., Shichijo S.;
RT "A new tumor-rejection antigen recognized by cytotoxic T lymphocytes
RT infiltrating into a lung adenocarcinoma.";
RL Cancer Res. 60:4830-4837(2000).

DR EMBL: AB018307; BAA34484.1; -;
DR EMBL: AF197954; AAG28523.1; -;
SQ SEQUENCE 414 AA; 46192 MW; 59724A96353D44D5 CRC64;

Query Match 100.0%; Score 37; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
DB 158 LYQAVATI 165

RESULT 4
Q9Z1A9 PRELIMINARY; PRT; 891 AA.
AC Q9Z1A9;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE BUB2-like protein 1 (Vascular RAB-GAP/TBC-containing).
GN VRP OR HBLP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEMATOPOIETIC;
RX MEDLINE=96017645; PubMed=7588067;
RA Guimaraes M.J., Bazan J.F., Zlotnik A., Wiles M.V., Grimaldi J.C.,
RA Lee F., McClanahan T.;
RT "A new approach to the study of haematopoietic development in the yolk
RT sac and embryoid bodies.";
RL Development 121:3335-3346(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=HEMATOPOIETIC;
RA Guimaraes M.J., Bazan J.F.;
RT "The CHES protein domain: a novel structural unit that is common
RT among networks involved in cell signaling, chromatin metabolism and
RT cell division.";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: U88873; AAD00658.1; -;
DR EMBL: BC005421; AAH05421.1; -;
DR MGD: MGI:1927225; Vrp.
DR InterPro: IPR004182; GRAM_dom.
DR InterPro: IPR000195; RabGAP_TBC.
DR Pfam: PF02893; GRAM; 1.
DR Pfam: PF00566; TBC; 1.
DR SMART: SM00164; TBC; 1.
SQ SEQUENCE 891 AA; 101976 MW; 394A12B904D06899 CRC64;

Query Match 94.6%; Score 35; DB 11; Length 891;
Best Local Similarity 75.0%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
DB 769 LYQAVATI 776

RESULT 5
O95759 PRELIMINARY; PRT; 897 AA.
AC O95759;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:46:58 ; Search time 1.74408 Seconds
(without alignments)
945.129 Million cell updates/sec

Title: US-09-857-308-5
Perfect score: 37
Sequence: 1 LYQAVATI 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-rvirus:*
16: sp-bacteriap:*
17: sp-archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	37	100.0	412	4	Q9H2T6
2	37	100.0	412	11	Q9CZV5
3	37	100.0	414	4	Q94864
4	35	94.6	891	11	Q921A9
5	35	94.6	897	4	Q95759
6	32	86.5	100	16	Q9A065
7	32	86.5	273	1	Q74043
8	32	86.5	485	17	Q8TL50
9	32	86.5	865	3	Q9U081
10	31	83.8	269	1	Q74066
11	30	81.1	129	15	Q36923
12	30	81.1	129	15	Q9QFQ8
13	30	81.1	129	15	Q9QFQ7
14	30	81.1	129	15	Q9QFQ1
15	30	81.1	130	15	Q36832
16	30	81.1	130	15	Q9QFQ2

17	30	81.1	130	15	Q9QFQ0	Q9qfq0	human	immun
18	30	81.1	130	15	Q9QFQ1	Q9qfk1	human	immun
19	30	81.1	130	15	Q9QFQ0	Q9qfk0	human	immun
20	30	81.1	130	15	Q905Y5	Q905y5	human	immun
21	30	81.1	130	15	Q905Y4	Q905y4	human	immun
22	30	81.1	130	15	Q905Y1	Q905y1	human	immun
23	30	81.1	131	15	Q9QFT7	Q9qft7	human	immun
24	30	81.1	131	15	Q9QFT6	Q9qft6	human	immun
25	30	81.1	131	15	Q9QFK2	Q9qfk2	human	immun
26	30	81.1	132	15	Q9QFQ3	Q9qfu3	human	immun
27	30	81.1	132	15	Q9QFU1	Q9qfu1	human	immun
28	30	81.1	132	15	Q9QFU0	Q9qfu0	human	immun
29	30	81.1	132	15	Q9QFT9	Q9qft9	human	immun
30	30	81.1	132	15	Q9QFT5	Q9qft5	human	immun
31	30	81.1	133	15	Q9QG05	Q9qg05	human	immun
32	30	81.1	133	15	Q9QG04	Q9qg04	human	immun
33	30	81.1	133	15	Q9QG03	Q9qg03	human	immun
34	30	81.1	133	15	Q9QG02	Q9qg02	human	immun
35	30	81.1	133	15	Q9QG01	Q9qg01	human	immun
36	30	81.1	133	15	Q9QFU4	Q9qfu4	human	immun
37	30	81.1	133	15	Q9QFU2	Q9qfu2	human	immun
38	30	81.1	133	15	Q9QFT8	Q9qft8	human	immun
39	30	81.1	133	15	Q9QFS4	Q9qfs4	human	immun
40	30	81.1	133	15	Q9QFS3	Q9qfs3	human	immun
41	30	81.1	133	15	Q9QFS2	Q9qfs2	human	immun
42	30	81.1	133	15	Q9QFS1	Q9qfs1	human	immun
43	30	81.1	133	15	Q9QFS0	Q9qfs0	human	immun
44	30	81.1	133	15	Q9QFR9	Q9qfr9	human	immun
45	30	81.1	133	15	Q9QFR8	Q9qfr8	human	immun

ALIGNMENTS

RESULT 1

ID Q9H2T6 PRELIMINARY; PRT; 412 AA.
AC Q9H2T6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE Antigen ART1/P17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ADENOCARCINOMA;
RX MEDLINE=20441578; PubMed=10987294;
RA Nishizaka S., Gomi S., Harada K., Oizumi K., Itoh K., Shichijo S.;
RT "A new tumor-rejection antigen recognized by cytotoxic T lymphocytes
infiltrating into a lung adenocarcinoma.";
RL Cancer Res. 60:4830-4837(2000).
DR EMBL: AF224759; AAG47636.1; -.
SQ SEQUENCE 412 AA; 45950 MW; 99A3830B2E4D3E22 CRC64;

Query Match 100.0%; Score 37; DB 4; Length 412;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

Db 156 LYQAVATI 163

RESULT 2

ID Q9CZV5 PRELIMINARY; PRT; 412 AA.

AC Q9CZV5;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

```

DE Vascular Rab-GAP/TBC-containing protein.
GN VRP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99303817; PubMed=10373574;
RX Yonekura H., Migita H., Sakurai S., Wang H., Harada S., Abedin M.J.,
RA Yamagishi S., Yamamoto H.;
RT "Antisense display-a method for functional gene screening: evaluation
RT in a cell-free system and isolation of angiogenesis-related genes.";
RL Nucleic Acids Res. 27:2591-2600(1999).
DR EMBL: AB024057; BAA75489.1; -
DR InterPro: IPR004182; GRAM.dom.
DR InterPro: IPR000195; RabGAP_TBC.
DR Pfam: PF02893; GRAM; 1.
DR Pfam: PF00566; TBC; 1.
DR SMART: SM00164; TBC; 1.
SQ SEQUENCE 897 AA; 102676 MW; D498C280A657BD1C CRC64;

Query Match 94.6%; Score 35; DB 4; Length 897;
Best Local Similarity 75.0%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 770 LYQAIATV 777

RESULT 6
ID Q9A065 PRELIMINARY; PRT; 100 AA.
AC Q9A065;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein SPY0914.
GN SPY0914.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296236;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: AE006940; AAK33830.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 100 AA; 12041 MW; F06A809D672C5D6C CRC64;

Query Match 86.5%; Score 32; DB 16; Length 100;
Best Local Similarity 87.5%; Pred. No. 8.5;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 76 LYQEAVTI 83

RESULT 7
ID O74043 PRELIMINARY; PRT; 273 AA.
AC O74043;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

```

```

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 31.1 kDa protein.
OS Cenarchaeum symbiosum.
OC Archaea; Crenarchaeota; Cenarchaeum.
OX NCBI_TaxID=46770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A;
RX MEDLINE=98422450; PubMed=9748430;
RA Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
RA Swanson R.V.;
RT "Genomic analysis reveals chromosomal variation in natural populations
RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";
RL J. Bacteriol. 180:5003-5009(1998).
DR EMBL: AF083071; AAC62686.1; -
DR InterPro: IPR003115; ParBc.
DR InterPro: IPR004437; ParB_part.
DR Pfam: PF02195; ParBc; 1.
DR SMART: SM00470; ParBc; 1.
DR TIGRFAMS: TIGR00180; parB_part; 1.
KW Hypothetical protein.
SQ SEQUENCE 273 AA; 31103 MW; 683C683F18DF61BE CRC64;

Query Match 86.5%; Score 32; DB 1; Length 273;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 178 LYQAVPTI 185

RESULT 8
ID Q8TLSO PRELIMINARY; PRT; 485 AA.
AC Q8TLSO;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Proton/sodium-glutamate symporter.
GN MA2961.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArelano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL: AE010998; AAM06334.1; -
KW Complete proteome.
SQ SEQUENCE 485 AA; 51158 MW; 4A3C504F95A1B322 CRC64;

Query Match 86.5%; Score 32; DB 17; Length 485;
Best Local Similarity 75.0%; Pred. No. 49;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 111

```

Db 348 LYQGVATV 355

RESULT 9

Q9U081 PRELIMINARY; PRT; 865 AA.

AC Q9U081; 174 LYQAVPTV 181

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Putative gamma-adaptin.

SPCPIEL1.06.

GN Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972H-;

RA Wedler H., Duesterhoeft A., McDougall R.C., Rajandream M.A.,

RA Barrell B.G.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL01183; CAB54865.1; -

DR InterPro; IPR002553; Adaptin_N.

DR InterPro; IPR001121; Gamma_adaptin_C.

DR Pfam; PF01602; Adaptin_N.1.

DR Pfam; PF02883; Alpha_adaptin_C2.1.

DR ProDom; PD021457; Gamma_adaptin_C.1.

SQ SEQUENCE 865 AA; 96041 MW; 05BA43A15B61E84E CRC64;

Query Match 86.5%; Score 32; DB 3; Length 865;

Best Local Similarity 87.5%; Pred. No. 92;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8

Db 317 LYQAVRTI 324

RESULT 10

O74066 PRELIMINARY; PRT; 269 AA.

AC O74066;

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Hypothetical 30.7 kDa protein.

OS Cenarchaeum symbiosum.

OC Archaea; Crenarchaeota; Cenarchaeum.

OX NCBI_TaxID=46770;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=B;

RX MEDLINE=98422450; PubMed=9748430;

RA Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,

RA Swanson R.V.;

RT "Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";

RL J. Bacteriol. 180:5003-5009(1998).

DR EMBL; AF083072; AAC62709.1; -

DR InterPro; IPR003115; ParBc.

DR InterPro; IPR004437; ParB_part.

DR Pfam; PF02195; ParBc.1.

DR SMART; SM00470; ParBc.1.

DR TIGRfams; TIGR00180; parB_part.1.

KW Hypothetical protein.

SQ SEQUENCE 269 AA; 30689 MW; AA45216CAD16DD5B CRC64;

Query Match 83.8%; Score 31; DB 1; Length 269;

Best Local Similarity 75.0%; Pred. No. 44;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8

Db 174 LYQAVPTV 181

RESULT 11

O36923 PRELIMINARY; PRT; 129 AA.

ID O36923

AC O36923;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE GAG polyprotein (Fragment).

GAG.

GN Human immunodeficiency virus type 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97445059; PubMed=9300048;

RA Leigh Brown A.J., Lobidel D., Wade C.M., Rebus S., Phillips N.,

RA Brettle R.P., France A.J., Leen C.S., McMenamin J., McMillan A.,

RA Maw R.D., Mulcahy F., Robertson J.R., Sankar K.N., Scott G., Wyld R.,

RA Peutherer J.F.;

RT "The molecular epidemiology of human immunodeficiency virus type 1 in six cities in Britain and Ireland.";

RL Virology 235:166-177(1997)

DR EMBL; AF014320; AAC58401.1; -

DR InterPro; IPR000071; Retrovir_p17.

DR Pfam; PF00540; Gag_p17.1.

KW AIDS; Core protein; Polyprotein.

FT NON_TER 1

FT NON_TER 129

SQ SEQUENCE 129 AA; 14468 MW; 22D4B58167285DF9 CRC64;

Query Match 81.1%; Score 30; DB 15; Length 129;

Best Local Similarity 75.0%; Pred. No. 34;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8

Db 55 LYNAVATL 62

RESULT 12

Q9QF08 PRELIMINARY; PRT; 129 AA.

ID Q9QF08

AC Q9QF08;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE GAG polyprotein (Fragment).

GAG.

GN Human immunodeficiency virus type 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LN23;

RX MEDLINE=99412391; PubMed=10482626;

RA Morris A., Marsden M., Halcrow K., Hughes E.S., Brettle R.P.,

RA Bell J.E., Simmonds P.;

RT "Mosaic structure of the human immunodeficiency virus type 1 genome infecting lymphoid cells and the brain: evidence for frequent in vivo recombination events in the evolution of regional populations.";

RL J. Virol. 73:8720-8731(1999).

DR EMBL; AF174940; AAF00286.1; -

DR InterPro; IPR000071; Retrovir_p17.

DR Pfam; PF00540; Gag_p17.1.

DR PRINTS; PR00234; HIV1MATRIX.

KW AIDS; Core protein; Polyprotein.

FT NON_TER 1

FT NON_TER 129


```
SQ SEQUENCE 129 AA; 14574 MW; C24B7A484562E0 CRC64;
Query Match 81.1%; Score 30; DB 15; Length 129;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 51 LYNNAVATL 58

RESULT 13
Q9QFQ7 PRELIMINARY; PRT; 129 AA.
AC Q9QFQ7;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE GAG polyprotein (Fragment).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LN24;
RX MEDLINE=99412391; PubMed=10482626;
RA Morris A., Marsden M., Halcrow K., Hughes E.S., Brettle R.P.,
RA Bell J.E., Simmonds P.;
RT "Mosaic structure of the human immunodeficiency virus type 1 genome
RT infecting lymphoid cells and the brain: evidence for frequent in vivo
RT recombination events in the evolution of regional populations.";
RL J. Virol. 73:8720-8731(1999).
DR EMBL: AF174941; AAF00287.1; -
DR InterPro: IPR000071; Retrovir_p17.
DR Pfam: PF00540; Gag_p17; 1.
DR PRINTS: PR00234; HIVMATRIX.
DR AIDS: Core protein; Polyprotein.
KW NON_TER 1
FT NON_TER 129
SQ SEQUENCE 129 AA; 14448 MW; C2234C7181D896B3 CRC64;

Query Match 81.1%; Score 30; DB 15; Length 129;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 51 LYNNAVATL 58

RESULT 14
Q9QFQ1 PRELIMINARY; PRT; 129 AA.
AC Q9QFQ1;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE GAG polyprotein (Fragment).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LN5;
RX MEDLINE=99412391; PubMed=10482626;
RA Morris A., Marsden M., Halcrow K., Hughes E.S., Brettle R.P.,
RA Bell J.E., Simmonds P.;
RT "Mosaic structure of the human immunodeficiency virus type 1 genome
RT infecting lymphoid cells and the brain: evidence for frequent in vivo
RT recombination events in the evolution of regional populations.";
RL J. Virol. 73:8720-8731(1999).
```

```
DR EMBL: AF174947; AAF00293.1; -
DR InterPro: IPR000071; Retrovir_p17.
DR Pfam: PF00540; Gag_p17; 1.
DR PRINTS: PR00234; HIVMATRIX.
KW AIDS: Core protein; Polyprotein.
FT NON_TER 1
FT NON_TER 129
SQ SEQUENCE 129 AA; 14474 MW; 007EF7119CFFEF41 CRC64;

Query Match 81.1%; Score 30; DB 15; Length 129;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 51 LYNNAVATL 58

RESULT 15
O36832 PRELIMINARY; PRT; 130 AA.
AC O36832;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE GAG polyprotein (Fragment).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97445059; PubMed=93000048;
RA Leigh Brown A.J., Lobidel D., Wade C.M., Rebus S., Phillips N.,
RA Brettle R.P., France A.J., Leen C.S., McMenamin J., McMillan A.,
RA Maw R.D., Mulcahy F., Robertson J.R., Sankar K.N., Scott G., Wyld R.,
RA Peutherer J.F.;
RT "The molecular epidemiology of human immunodeficiency virus type 1 in
RT six cities in Britain and Ireland.";
RL Virology 235:166-177(1997).
DR EMBL: AF014229; AAC58310.1; -
DR InterPro: IPR000071; Retrovir_p17.
DR Pfam: PF00540; Gag_p17; 1.
DR PRINTS: PR00234; HIVMATRIX.
KW AIDS: Core protein; Polyprotein.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 14450 MW; E7AA32546F1B2656 CRC64;

Query Match 81.1%; Score 30; DB 15; Length 130;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 55 LYNNAVATL 62

Search completed: July 18, 2003, 09:59:38
Job time : 2.74408 secs
```

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:46:58 ; Search time 90.2559 Seconds
(without alignments)
945.129 Million cell updates/sec

Title: US-09-857-308-1
Perfect score: 2175
Sequence: 1 MNLQRYWGEIPISSTQNRSS.....SLMGSSPVFNQCKRMRKI 414

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2175	100.0	414	4	Q94864
2	2155	99.1	412	4	Q9H2T6
3	2079	95.6	412	11	Q9CZV5
4	142.5	6.6	979	3	P87152
5	132	6.1	359	5	Q9VX12
6	131	6.0	1049	5	Q9V149
7	126	5.8	1621	16	Q9KTA5
8	121.5	5.6	505	3	Q8XOV8
9	121.5	5.6	2957	5	O61845
10	121.5	5.6	2977	5	Q9VAP9
11	120.5	5.5	893	11	Q99PI5
12	118.5	5.4	582	11	Q9ER99
13	117	5.4	335	4	Q9Y4V2
14	117	5.4	486	4	Q96HV9
15	116.5	5.4	791	11	Q9Z1P7
16	114.5	5.3	17352	5	Q95YM2

17	114	5.2	486	11	Q9QV18
18	113	5.2	1152	4	Q92603
19	113	5.2	1204	4	O00211
20	113	5.2	1235	4	Q9H2G2
21	113	5.2	1852	5	Q9VH42
22	112	5.1	486	4	Q9UNF0
23	112	5.1	1278	4	Q9UPP5
24	111.5	5.1	660	5	Q9VT63
25	111.5	5.1	1248	3	Q8XOV7
26	109.5	5.0	6815	5	Q917U4
27	109.5	5.0	16215	5	Q9NFS3
28	109	5.0	486	11	Q9WVE8
29	109	5.0	678	10	Q94I68
30	108.5	5.0	632	4	O94937
31	108.5	5.0	1206	11	O08815
32	108.5	5.0	1700	5	Q9VVA9
33	108	5.0	1266	5	Q9GRH4
34	107.5	4.9	568	12	Q9DUE2
35	107.5	4.9	3263	5	Q917U3
36	107	4.9	884	5	O61543
37	107	4.9	884	5	Q9VEV9
38	107	4.9	2484	5	Q9U347
39	107	4.9	2607	5	Q23187
40	106	4.9	599	10	Q93ZU5
41	106	4.9	599	10	Q9LRR7
42	106	4.9	611	5	O16938
43	106	4.9	1180	5	O18673
44	106	4.9	4667	5	Q9TVI9
45	106	4.9	5170	5	Q17490

ALIGNMENTS

RESULT 1

ID	Q94864	PRELIMINARY;	PRT;	414 AA.
AC	Q94864;			
DT	01-MAY-1999 (TrEMBLrel. 10, Created)			
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)			
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)			
DE	KIAA0764 protein (Adenocarcinoma antigen ART1).			
GN	KIAA0764.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRAIN;			
RX	MEDLINE=99087487; PubMed=9872452;			
RA	Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,			
RA	Kotani H., Nomura N., Ohara O.;			
RT	"Prediction of the coding sequences of unidentified human genes. XI.			
RT	The complete sequences of 100 new cDNA clones from brain which code			
RL	for large proteins in vitro."			
RL	DNA Res. 5:277-286(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20441578; PubMed=10987294;			
RA	Nishizaka S., Gomi S., Harada K., Oizumi K., Itoh K., Shichiho S.;			
RT	"A new tumor-rejection antigen recognized by cytotoxic T lymphocytes			
RT	infiltrating into a lung adenocarcinoma."			
RL	Cancer Res. 60:4830-4837(2000).			
DR	EMBL; AB018307; BAA34484.1;			
DR	EMBL; AF197954; AAC28523.1;			
SQ	SEQUENCE 414 AA; 46192 MW; 59724A96353D44D5 CRC64;			

Query Match 100.0%; Score 2175; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.8e-166;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNLQRYWGEIPISSTQNRSSFDLPREFRLVEVHDPPLHQPSSANKPKPTMLDIPSEPC 60

Db 1 MNLQRYWGEIPISSQTNRSFDDLPRFRLVEVDPLHQPSSANKPKPTMLDIPSEPC 60
Qy 61 SLTHTIQLIQRNRLNLIAQAQNOOQTEGVKTESEPLPSCPGSPPLPDDLPLDC 120
Db 61 SLTHTIQLIQRNRLNLIAQAQNOOQTEGVKTESEPLPSCPGSPPLPDDLPLDC 120
Qy 121 KNPAPFOIRHSDPESDFYRGKGEFVTELSWHSCROLLYQAVATILAHAGFDCANESYLE 180
Db 121 KNPAPFOIRHSDPESDFYRGKGEFVTELSWHSCROLLYQAVATILAHAGFDCANESYLE 180
Qy 181 TLTDVAHEYCKLFTKLLRFVADREARLGQTTPDVMQOVFHEVGIGSVLSLQKFWQHRKI 240
Db 181 TLTDVAHEYCKLFTKLLRFVADREARLGQTTPDVMQOVFHEVGIGSVLSLQKFWQHRKI 240
Qy 241 DYHSYMLQISKOLSEERYIVNPEKATEDAKPVKIKEEPVSDITFPVSEELASDQSLPM 300
Db 241 DYHSYMLQISKOLSEERYIVNPEKATEDAKPVKIKEEPVSDITFPVSEELASDQSLPM 300
Qy 301 SLPNGVLGAQSERFPPSNLEVEASQAASSAEVNASPLNLAHVKNMPEQSEEGNYSVGHVL 360
Db 301 SLPNGVLGAQSERFPPSNLEVEASQAASSAEVNASPLNLAHVKNMPEQSEEGNYSVGHVL 360
Qy 361 GSDVFEPMGSMGSEAGIPQSPDDSDSSYGSHTDSLGMSSPVFNQCKKMRKI 414
Db 361 GSDVFEPMGSMGSEAGIPQSPDDSDSSYGSHTDSLGMSSPVFNQCKKMRKI 414

RESULT 2
G9H2T6

Db 243 YMLQISKOLSEERYIVNPEKATEDAKPVKIKEEPVSDITFPVSEELASDQSLPM 302
Qy 305 GVLGAQSERFPPSNLEVEASQAASSAEVNASPLNLAHVKNMPEQSEEGNYSVGHVLGSDV 364
Db 303 GVLGAQSERFPPSNLEVEASQAASSAEVNASPLNLAHVKNMPEQSEEGNYSVGHVLGSDV 362
Qy 365 FEPMGSMGSEAGIPQSPDDSDSSYGSHTDSLGMSSPVFNQCKKMRKI 414
Db 363 FEPMGSMGSEAGIPQSPDDSDSSYGSHTDSLGMSSPVFNQCKKMRKI 412

RESULT 3
Q9CZV5
ID Q9CZV5 PRELIMINARY; PRT; 412 AA.
AC Q9CZV5;
DT 01-JUN-2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE 2610524B01RIK protein.
GN 2610524B01RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK012134; BAB28053.1; -;
DR MGD; MGI:1919445; 2610524B01RIK.
SQ SEQUENCE 412 AA; 45970 MW; B7BF1525BB44F6D5 CRC64;

Query Match 95.68; Score 2079; DB 11; Length 412;
Best Local Similarity 95.9%; Pred. No. 9.2e-159;
Matches 393; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Qy 5 RYWGEPISQTNRSFDDLPRFRLVEVDPLHQPSSANKPKPTMLDIPSEPCSLTI 64
Db 3 RYWGEPISQTNRSFDDLPRFRLVEVDPLHQPSSANKPKPTMLDIPSEPCSLTI 62
Qy 65 HTTOLIOHNRRLNLIAQAQNOOQTEGVKTESEPLPSCPGSPPLPDDLPLDCKNPN 124
Db 63 HTTOLIOHNRRLNLIAQAQNOOQTEGVKTESEPLPSCPGSPPLPDDLPLDCKNPN 122
Qy 125 APFOIRHSDPESDFYRGKGEFVTELSWHSCROLLYQAVATILAHAGFDCANESVLETLTD 184
Db 123 APFOIRHSDPESDFYRGKGEFVTELSWHSCROLLYQAVATILAHAGFDCANESVLETLTD 182
Qy 185 VAHEYCKLFTKLLRFVADREARLGQTTPDVMQOVFHEVGIGSVLSLQKFWQHRKIDYHS 244
Db 183 VAHEYCKLFTKLLRFVADREARLGQTTPDVMQOVFHEVGIGSVLSLQKFWQHRKIDYHT 242
Qy 245 YMLQISKOLSEERYIVNPEKATEDAKPVKIKEEPVSDITFPVSEELASDQSLPM 304

Db 63 HTTOLIOHNRRLNLIAQAQNOOQTEGVKTESEPLPSCPGSPPLPDDLPLDCKNPN 122
Qy 125 APFOIRHSDPESDFYRGKGEFVTELSWHSCROLLYQAVATILAHAGFDCANESVLETLTD 184
Db 123 APFOIRHSDPESDFYRGKGEFVTELSWHSCROLLYQAVATILAHAGFDCANESVLETLTD 182
Qy 185 VAHEYCKLFTKLLRFVADREARLGQTTPDVMQOVFHEVGIGSVLSLQKFWQHRKIDYHS 244
Db 183 VAHEYCKLFTKLLRFVADREARLGQTTPDVMQOVFHEVGIGSVLSLQKFWQHRKIDYHS 242
Qy 245 YMLQISKOLSEERYIVNPEKATEDAKPVKIKEEPVSDITFPVSEELASDQSLPM 304

```
Db 243 YMLQISQKSEEEYRINPKATEDTKPVKIKEEVPVSDITFPVSEELADLASQSLPI 302
Qy 305 GVLGAQSERFNSLEVEASQASSAEVNASPLNLAHVKMEPOSEBGNVGHVGLGSDV 364
Db 303 GVLGAQSERFNSLEVEASQASSAEVNASPLNLAHVKMEPOSEBGNVSAHVGLGSDV 362
Qy 365 FEPMSCMSBAGIQSPDDSDSSYGHSHSTDSLMGSSPVFNQRCCKMRKI 414
Db 363 FEPMSCMSBAGLQSPDDSDSSYGHSHSTDSLMGSSPVFNQRCCKMRKI 412

RESULT 4
P87152 PRELIMINARY; PRT; 979 AA.
ID P87152
AC P87152;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative transcriptional activator C25H2.11C.
GN SPBC25H2.11C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Darso G., Lye G., Bowman S., Church C., Wood V., Barrell B.G.,
RA Rajadream M.A., Connor R.E.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS A COPY OF THE BROMODOMAIN.
DR EMBL: 295397; CAB08777.1; -.
DR HSP: Q92831; I991.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; Bromodomain; 1.
DR SMART: SM00297; BROMO: 1.
DR PROSITE: PS00633; BROMODOMAIN.1; 1.
DR PROSITE: PS00014; BROMODOMAIN.2; 1.
KW Hypothetical protein; Transcription regulation; Nuclear protein;
KW Activator; Bromodomain.
FT DOMAIN 316 386 BROMODOMAIN (BY SIMILARITY).
SQ SEQUENCE 979 AA; 111503 MW; D60E78663B8C68AF CRC64;

Query Match 6.6%; Score 142.5; DB 3; Length 979;
Best Local Similarity 19.4%; Pred. NO. 0.011;
Matches 83; Conservative 74; Mismatches 183; Indels 87; Gaps 14;

Qy 10 IPTSSQT--NRSSFLLPREFLVEHPDPLHOPSKPKPTMLDIPSEPCSLTIHTI 67
Db 596 VPMTAGSVLESEEDLYFRDYSLFEI-----NRNTPGV-----SLMKNI 637
Qy 68 QLTQHNRLNLNLIATAQNOQTEGVKTESEPLESCPSPLPDLPLCKNPNAF 127
Db 638 AKMQEIRKLCNKIQTRQLQLPQPFYEHKSH-VFPANNEPILLD--IPQNYDNMSSEK 694
Qy 128 QIRHSDPESDFYKGEKPEPTELSWHSCROLLYQAVATILAHAGFCANSVLETLTDVAH 187
Db 695 PLAH-----DVLKLCILILFHAGFESFGMALDALTETIAA 730
Qy 188 EYCLK----FTKLLREAVDREARLGOTPPFDVMEQVFEHVGISVLSLQKFMWHRKDYH 243
Db 731 DYMAKGAAMDQYLIYKDKSQO-----EIVGQTLGELGVDDVNDLISYVYHVEROS 783
Qy 244 SYMLQISQKSEEEYRINPKATEDAKPVKIKEEVPVSDITFPVSE-----L 291
Db 784 VKLLETHQRLQRHFVLLRPAISERNDDEAIFNQGESFVTGNFVETGDDFFGLRELGL 843
Qy 292 EADLASGDSLPMGVLCGASERPPSNLEVEASQAS--SAEVNASPLNLAHVKMEPOSE 350
Db 844 DRELGLDLSLVLHLL---QSLRSLNKSQW--PEATIKGDQEVAPPKPPITAESISNE 898
```

```
Qy 351 EGNVSGHVGVLGSDVF-----EPPMSGMSEAGIQSPDDSDSSYGHSHSTDSLMGSSPVF 403
Db 899 IGLIQGLFKLNLEEFGLDELLEDEDIRPSKPPRPLPNGKITTCR----KRIASSVFL 954
Qy 404 NQRCKKR 410
Db 955 NQSLRKK 961

RESULT 5
Q9VX12 PRELIMINARY; PRT; 359 AA.
ID Q9VX12
AC Q9VX12;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG6506 protein.
GN CG6506.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003506; AAF48771.1; -.
DR FlyBase: FBgn0030874; CG6506.
SQ SEQUENCE 359 AA; 39460 MW; 26440857AFAD6854 CRC64;
```

Query Match 6.1%; Score 132; DB 5; Length 359;
Best Local Similarity 18.7%; Pred. NO. 0.018;
Matches 71; Conservative 71; Mismatches 174; Indels 64; Gaps 12;


```
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL: AE004181; AAF94159.1; -.
KW TIGR; VC0998; -.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 1621 AA; 178073 MW; 2EF38BAC27A2485D CRC64;

Query Match          5.8%; Score 126; DB 16; Length 1621;
Best Local Similarity 21.7%; Pred. No. 0.46;
Matches 93; Conservative 75; Mismatches 169; Indels 92; Gaps 20;

QY 9 EIPSSQTNRSFDLL-----PREFRLVEHDPPLHQPSSANKPKPPTMLDIPSPCSLT 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 815 DLPEQTATNETAFDELLADLAAPQSNVTDTSDAL-----APDGLSQSVEEP--LT 864
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 IHTIQLIQRNRLNL-IATAQAQNOQQTGKYTESEPLPSCPGSP-----PLP 112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 865 LNDLELPENDPEQLAEVTPSSAFDEQQVETEIEPESEPLAAEASNDSDLTALNELDLP 924
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 113 ----DDLPLDCKNNAFQIRHSDPESDFYRGKGPVTELSHSCRLLYQA--TTL 166
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 925 EYTEEDVLADVQLEPAEAEV---EPDLELV---NEPVTAEAFTELDLDPYTEEDAL 978
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 167 AHAGFDCAVESVLETLTDAHEYCLKFKLLRFAVDREA--RLGQTPTPDVMEQVFEHVG 224
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 979 ADAQLEPAEVAEVEPELDLASE-----PAEEAEFTLNLKLDLPYTEE----- 1021
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 225 IGSVLSLOKFWQHRKIDYHSYMLQISKOLSE-----EYERIVNPEKATEDA----- 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1022 --DALADQALESATESEVESELELVSEPAEAEAFTELDLDPYTEEDALADSOLEPAA 1079
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 271 -----KPKVKEEVPDITFPVSELEADLASGQSLPMGVL--GAQSERFPSNLEVAS 323
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1080 ESEVEPELSEVPEETEAFTLDELDPYTEEDALADAQLEPAVESEVEP-ELELASE 1138
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 324 P--QASSAEVNASPLWN-----LAHVMEPESEEGNVSGHVLGSDV-FEPMMSGNSE 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1139 PAEEEAETLNDLDPYTEEDALADAQLEP--AAEVESESELELASDLEEEPEFTLNE 1196
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 375 AGIQSPSD 383
   : : : :
DB 1197 LDLPEYTEE 1205

RESULT 8
Q8X0V8 PRELIMINARY; PRT; 505 AA.
AC Q8X0V8
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical 55.6 kDa protein.
GN 123A4.340.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Align V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL670009; CAD21385.1; -.
KW Hypothetical protein.

SQ SEQUENCE 505 AA; 55613 MW; 4E88D0BBA42383E7 CRC64;

Query Match          5.6%; Score 121.5; DB 3; Length 505;
Best Local Similarity 17.7%; Pred. No. 0.2;
Matches 76; Conservative 69; Mismatches 184; Indels 101; Gaps 16;

QY 10 IPISSQTNRSFDLLPREFRLVEHDPPLHQPSSANKPKPPTMLDIPSPCSLTHTITQL 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 22 VPLPSYADNSSTMSSSGRLRSRRPAPPSPPPSPKSRKATSSRAAPEPP-----QETVQP 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70 IQHNRRLNLATAQAQNOQQTGKYTESEPL---PSCPG--SPPL-PDDLPLDCKNP 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 78 V-----SPPLPHEDEEQPKATPKQMPILRPPTQTGTSPPPQHPDDIYRMPPLPP 126
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 124 NAFQIRHSDPESDFYRGKGPVTELSHSCRLLYQA--VATILAHAGFDCAVESVLETL 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 127 YLSVQPRICEP-ADNYRG-----WTTPLFESYKMNHPLTYGSF----- 165
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 TDVAHEYCLKFKLLRFAVDREARLGQTPFPDME-----QVFHEVGIGS--- 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 166 -DIWFEQVQKMAEASNCSELTNTRICKPDPGTVAEAFICQMRFLSCWRILHGTISGAYWA 224
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 228 ---VLSLOKF-----WQHRKIDYHSYMLQISKOL---SEEVERIVN 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 225 YNRVLGYSRIPVFTGAPDGGSGEQSGACQWQPSPDCYVFALEVKRVTPVNSEQPPWYQ 284
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 263 PEKATEDAKPVKIKEP-----VSDITFPVSELEADLASGQSLPMGVLGAQSERFPSNL 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 285 TNRWVEEVSRAKVTIYPSVQYKEGMWLKEVNLGMILKGDQVCESMYDRKPDWAPWP 344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 319 EYEASQASSAEVNASPLWNLAHVMEPESEEGNVSGHVLGSDVFEPMMSGNSEAGIP 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 345 EVPAGGE-----VEPPWALKGSPLRAEDAFSGSGAGVYVSS-----KNAGHIGQQ 391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 379 QSPDSDSY 388
   : : : :
DB 392 HOPQQAQSTF 401

RESULT 9
O61845 PRELIMINARY; PRT; 2957 AA.
ID O61845
AC O61845;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE T04D1.4 protein.
GN T04D1.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RC MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldmann P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Davidson S., Wohldmann P.;
```



```
Db 2230 RQNGKTFQ-----ILMQESTLPHRT-----SELLQMOQLMIDDLKGEINEIRTAKELEQ 2280
Qy 274 KIKEEPYSDITFPVSEELADSLGSDSLPMGLVGAQSERPPSNLEVEASPOA----- 326
Db 2281 QLKKEEQTSE-----KSTAESDVLSEAK-----VKEIKQEPIDKSAQDGVVDFDL 2327
Qy 327 --SSAEYNASPLMNLNAHVHAKPEQBS-----EEGNVSGHGVLSGVDFPEPMSCMSEAGTIPQ 379
Db 2328 LKNSADVKN-----AFSKSQPQQTIVTKQESNESGEPSPVSGVSKSEGHDKESKYLKK 2381
Qy 380 SPDDSD 385
Db 2382 DYDNFD 2387

RESULT 11
Q99P15 ID Q99P15 PRELIMINARY; PRT; 893 AA.
AC Q99P15;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE LPIN2.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20578762; PubMed=11138012;
RA Peterfy M., Phan J., Xu P., Reue K.;
RT "Lipodystrophy in the fld mouse results from mutation of a new gene
RT encoding a nuclear protein, lipin.";
RL Nat. Genet. 27:121-124(2001).
DR EMBL; AF286723; AAG52761.1;
DR MGI; MGI:1891341; Lpin2.
DR InterPro; IPR001395; Aldo/ket_red.
DR PROSITE; PS00063; ALDO-KETO-REDUCTASE_3; UNKNOWN_1.
SQ SEQUENCE 893 AA; 99617 MW; 3ADC0C24062B574 CRC64;

Query Match 5.5%; Score 120.5; DB 11; Length 893;
Best Local Similarity 18.9%; Pred. No. 0.54;
Matches 76; Conservative 52; Mismatches 136; Indels 139; Gaps 16;

Qy 7 WGEIPISSQTNRSSFDLLPR-----EFLVEVHDPPHLQPSAN----- 45
Db 266 WGGFPESTKVKTRYDYHPTATITPSENTHFRVIPSEDSLIREVEKDATVEDTCTIV 325
Qy 46 KKPPTM-----LDIPSEPCSLTIHTLIQHNRRLNLIATAQAQNOQOQTEGKVT 96
Db 326 KPRPRALCKQLSDAASTELPESP-----LEAPQISLLDADPVPS 365
Qy 97 EE-----SEPLPSCPGSPPL-----PDDLPLDCK--NPN-APFQIRHS----- 133
Db 366 PSAEAPSEPKPAKDSPTKKXGVHKSQHQPDDIYLLDLKALEPEVAALYFPKSDTPDG 425
Qy 134 ----PESDFRGKGEPTVELSWHSCROLLYQAVATILAHAGFCANESVLETLTDVAHEY 189
Db 426 SROMPESDTFSGSQSP-----QSVGSAADSGTECLSDSAMD-LPDVFTLSL 470
Qy 190 C-----LKFTKLLRFADVREARLGOTFPFDVMEQVFHE-----VGIGSVLSLQK 233
Db 471 CGGLSENGEISKEKFEHIIITYHEFAENPGLINPNLVIIRYNNYAWALAAPMILSLQV 530
Qy 234 FWQHRKIDYHMYLQISQKSEYERIVNPEKA-----TEDAKPVKIKEE 278
Db 531 FQK-----SLPKATVSWVKDKMPKSGRWFWFKKESMIKQLPETKEGKSEVP 579
Qy 279 PVSDITFPVSEELADSLGSDQLPMGLVGAQSERPPSNLEVE 321
Db 580 PANDLPSNABEPTSARPAENDTSSDEG-----SQELEFESIKVD 617
```

RESULT 12

```
Q9ER99 ID Q9ER99 PRELIMINARY; PRT; 582 AA.
AC Q9ER99;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Regulatory subunit of SGLT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Baumgarten K., Gorboulev V., Koepsell H.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y11917; CAA72676.1;
DR InterPro; IPR000449; UBA_domain.
DR SMART; SM00165; UBA; 1.
SQ SEQUENCE 582 AA; 61247 MW; 23CB8C50CA2B390F CRC64;

Query Match 5.4%; Score 118.5; DB 11; Length 582;
Best Local Similarity 20.8%; Pred. No. 0.43;
Matches 83; Conservative 52; Mismatches 162; Indels 103; Gaps 18;

Qy 40 HQPSANKPKPTMLDIPSEPCSLTIHTLIQHNRRLNLIATAQAQNOQO----- 90
Db 212 HLPSEAKGLPAS--GLCSCPCSEALMEVDTAEE-----QSLVAMCSSTGRQDAVIKPSVA 264
Qy 91 --TEGVKTERSEPLSCPGSPPLDPLLPLDCKNPNAPFOI-----RHSDPESDFYRGKE 144
Db 265 HLASDNTMETVETLQSNPCEPVEHSILTRQLQLPEDNVDMSTMDNDDSSLLSGHQ 324
Qy 145 PVTELSWHSCROLLYQAVATILAHAGFCANESVLETLTDVAHEYCKLFTKLLRFADVRE 204
Db 325 PSVESAEFFC-----SSVTVALKEL-HELLVTSCK----- 353
Qy 205 ARLGOTFPFDVMEQVFH-----EVGIGSVLSLQKFWHQHRIKDYH-----SYMLQISQKLS 254
Db 354 -----PASESPEHVTCQSEIGAESQPSVDSLGRVRVQSVHLTPSDQYSGSCHOAT 405
Qy 255 EEYER--IVN--PEKATEDAKPVKIKEEPVSDITFPVSEELADSLGSDQLPMGLVGAQ 310
Db 406 SESKTIIVGTAPCAAVEDEASTSF--EGLGDGLSPDREDVRRSTESARKSCSVAITSAK 463
Qy 311 -SERFPNLEVEASPOQSSAE-VNASPLMNLNAHV-KMPEOESBEGNVSGHGVLSGVDFEE 367
Db 464 LSEQLPCTGLVETAPELAASEGASQSP---SEHVHNPGRDPETSSVC----- 508
Qy 368 PMSCMSEAGIPQSPDDSDSS-----YGSHTSDLSLMS 399
Db 509 PGAGLPRSLGDDQPPTQSLSTPVSPLPPFIFFPAADVDRILGA 548

RESULT 13
Q9Y4V2 ID Q9Y4V2 PRELIMINARY; PRT; 335 AA.
AC Q9Y4V2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE D437M21.3 (Protein kinase C and casein kinase substrate in neurons 2)
DE (Fragment).
GN PACSIN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Laird G.;
```


Search completed: July 18, 2003, 09:59:37
Job time : 96.2559 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2003, 07:40:17 ; Search time 2561 Seconds
(without alignments)
10820.179 Million cell updates/sec

Title: US-09-857-308-2

Perfect score: 1711

Sequence: 1 acgcgatccttgcctcaggc.....aaaaaaaaaaaaaaaaaaaaa 1711

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1057	61.8	2612	11 AK012134	Mus muscu
2	940.4	55.0	1027	14 BM921322	AGENCOURT
3	904.6	52.9	939	13 BM545491	AGENCOURT
4	824.4	48.2	865	14 BQ229669	AGENCOURT
5	817.4	47.8	1214	13 BM545988	AGENCOURT
6	797.4	46.6	923	14 BQ924589	AGENCOURT

7	758.4	44.3	904	13	BI822539	603035248
8	750.6	43.9	818	13	BI659780	603293356
9	738.2	43.1	972	13	BM468753	AGENCOURT
10	731.2	42.7	938	12	EG681414	602627777
11	721.4	42.2	771	13	BI770385	603055622
12	710.8	41.5	735	14	BQ576180	UI-H-E21-
13	706	41.3	821	12	BG489862	602518753
14	704.4	41.2	718	14	BM728984	UI-E-E01-
15	702	41.0	899	14	BQ887284	AGENCOURT
16	700.6	40.9	788	13	BI763619	603050193
17	691.8	40.4	764	13	BI855674	603383202
18	686.4	40.1	798	13	BI193419	602947051
19	684.2	40.0	725	12	EG149842	nae01c08.
20	677.8	39.6	681	13	BI160260	602863964
21	665.8	38.9	758	12	EG706749	602671875
22	664.4	38.8	712	14	BM989602	UI-H-DH0-
23	658.6	38.5	796	9	AL559070	AL559070
24	656	38.3	783	13	BI602617	603248445
25	654.4	38.2	1059	12	EG323193	602421403
26	634.6	37.1	1105	14	BM802810	AGENCOURT
27	627.8	36.7	687	13	BM050170	603632231
28	626.4	36.6	1059	14	BM802006	AGENCOURT
29	615	35.9	616	9	AI130886	qb81e05.x
30	606	35.4	839	12	EG163932	602343111
31	601.8	35.2	688	13	EG927573	HNC43-1-C
32	596	34.8	608	12	EG776075	602663014
33	594	34.7	594	9	AA694172	z142a01.s
34	594	34.7	598	13	BM310780	BM310780
35	592.8	34.6	619	13	BI907412	iq48dl1.y
36	592.8	34.6	679	13	EG927465	HNC54-1-A
37	588	34.4	927	12	EG033451	602301836
38	586.6	34.3	624	10	BE280681	601155947
39	585.4	34.2	636	14	BM748747	K-EST0023
40	584.6	34.2	842	9	AL580770	AL580770
41	580.6	33.9	904	10	BE621165	601494101
42	580	33.9	649	12	BF947767	MR3-NN022
43	580	33.9	824	12	EG106653	602290482
44	579.4	33.9	787	12	EG705505	602688609
45	578	33.8	621	10	BE292913	601105354

ALIGNMENTS

RESULT 1	AK012134	2612 bp	linear	HTC 19-JAN-2002
AK012134	Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.			
LOCUS	AK012134			
DEFINITION	AK012134.1 GI:12848691			
ACCESSION	HTC; CAP trapper.			
VERSION	Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library			
KEYWORDS	clone:2610524B01.			
SOURCE	Mus musculus			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1			
AUTHORS	Carninci,P. and Hayashizaki,Y.			
TITLE	High-efficiency full-length cDNA cloning			
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)			
MEDLINE	99279253			
PUBMED	10349636			
REFERENCE	2			
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.			
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)			
MEDLINE	20499374			
PUBMED	11042159			

REFERENCE
AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazawa, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multipillar sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, F., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaudo, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marchionni, L., Washima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Schonwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schenbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyo-Oka, K., Wang, K.H., Weitz, C., Wittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.,
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085860
11217851

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

5 (bases 1 to 2612)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C.,
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Horii, F.,
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S.,
Kumihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
Numazaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamana, I.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.
Direct Submission

TITLE
JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>, Tel.81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.

FEATURES
source

Location/Qualifiers
1. 2612
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:2610524B01"
/db_xref="MGI:1904965"
/db_xref="taxon:10090"
/clone="2610524B01"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days embryo"
290. .1528

CDS

/note="data source:SPT, source key:Q9H2T6, evidence:ISS
homolog to ANTIGEN ART1/P17
putative"
/codon_start=1
/protein_id="BAB28053.1"
/db_xref="GI:12848692"
/translation="MLRYWGEIPISQGTNRSSFDLLPREFLVEVHPPHLPQHSANK
PKPTMLDIPSECSLTHTLIQHNRRLSLTAQTQSOQOETGVKABESEPLPS
CPGPPPLDDLPDCKNPAPFOIRSDPSDFVRGKEPVTLELGSWSHCRLLYQAV
ATILAHGCFECANESVLETLTDVAHEYCLKFTKLLRFADREALLGOTPFDDVMSOVE
HEVIGSVLSQLEKFWQHRIDKYHYMLQISQLSEERYINPKATDTPKVKKEE
PVSDITFPVSELEADLASQSLPIGVLGASDFPSNLEVASPPQAPAEKNSPL
SLMHAHVMEPQESNGNVAHVLGSLDVFEEPMSCMSAELGPQSPDDSDSYSGSHSTD
WLGSSPVFNQRCRMRK1"
2598. 2603
/note="putative"
2612
/note="putative"
BASE COUNT 641 a 687 c 622 g 662 t
ORIGIN
Query Match 61.8%; Score 1057; DB 11; Length 2612;
Best Local Similarity 90.4%; Pred. No. 3.2e-183;
Matches 1129; Conservative 0; Mismatches 120; Indels 0; Gaps 0;
Qy 344 AATCTGAAAGATCTACTGGGAGAGATACCAATATCATCAAGCCAGACCAAGAGAGTTCC 403
Db 287 AAATGCTGAGGTACTGGGAGAGATACCAATACCGTCAGCAGACCAAGAGAGTTCC 346
Qy 404 TTCGATTTGCTCCACGGGAGTTCGCTGCTGGTGAAGTCCATGACCCACCCCTGCACCAA 463
Db 347 TTTGATCTGCTCCACGGGAGTTCGCTGCTGGTGAAGTCCATGACCCACCCCTGCACCAA 406
Qy 464 CCCTCAGCCACAGCCGAGCCGCCACCTACTGCTGGACATCCCTCAGAGCCAGTGTAGT 523
Db 407 CCCTCAGCCACAGCCGAGCCGCCACCTACTGCTGGACATCCCTCAGAGCCGAGTGTAGT 466
Qy 524 CTCACCATCCATGATTCAGTTGATTCAGCAGCAACACGAGCTCTTCGCAACCTTATTGCC 583
Db 467 CTCACCATCCATGATTCAGTTGATTCAGCAGCAACACGAGCTCTTCGCAACCTTATTGCC 526
Qy 584 ACAGTCAGGCCAGAGATTCAGCAGCAGCAGAGAGTGTAAACACTGAAGAGAGTGAACCT 643
Db 527 ACAGTCAGGCCAGAGTTCAGCAGCAGCAGAGAGTGTAAAGGCTGAAGAGAGTGAACCT 586
Qy 644 CTTCCCTGCTGCTGGTGCACCTCTCTCTCTGATGACCTCTCTGCTGATGAGTGTAGTGAAG 703
Db 587 CTTCCCTGCTGCTGGTGCACCTCTCTCTCTGATGACCTCTCTGCTGATGAGTGTAGTGAAG 646
Qy 704 AATCCCAATGCACCATTCAGATCCGCGCAGCAGTGCACCCAGAGAGTGAAGTGTAGTGGG 763
Db 647 AATCCCAATGCACCATTCAGATCCGCGCAGCAGTGCACCCAGAGAGTGAAGTGTAGTGGG 706
Qy 764 AAAGGGGAACCTGTGAGTGAAGTGCAGTGGGACCTCTCTGCGGAGCTCTCTACCAAGCA 823
Db 707 AAAGGAGAGCCTGTGACAGAGTGTGAGTGGGACCTCTCTGCGGAGCTCTCTATCAAGCA 766
Qy 824 GTGGCCCAATCTGTGGCCCGCGGCTTTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGAGAC 883
Db 767 GTGGCCCAATCTGTGGCCCGCGGCTTTCAGTGTGCTGCTGCTGCTGCTGCTGCTGAGAC 826
Qy 884 CTAAGTGTGCGCAGATGAGTATTGCTTAAGTTTACCAAGTGTGCTGCTGCTGCTGCTGCTG 943
Db 827 CTAAGTGTGCGCAGATGAGTATTGCTTAAGTTTACCAAGTGTGCTGCTGCTGCTGCTGCTG 886
Qy 944 GACCGGAGGCCCGGTGGGACAGCTCTCTTTCTGTGATGAGGAGGAGTGTATTCAT 1003
Db 887 GACCGGAGGCCCGGTGGGACAGCTCTCTTTCTGTGATGAGGAGGAGTGTATTCAT 946
Qy 1004 GAAGTGGTATGTCAGTGTCTCTCTCTCCAGAGTGTGCGCAGCAGCCGATCAAGGAC 1063
Db 947 GAAGTGGTATGTCAGTGTCTCTCTCTCCAGAGTGTGCGCAGCAGCCGATCAAGGAC 1006

Qy	1064	TATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTGTC	1123
Db	1007	TATCACACTTACATGCTACAGATTAGTAAGCAGCTCTCTGAAGAATATGAAGGATTGTC	1066
Qy	1124	AATCCTGAGAGGCCACAGAGGAGCCTAAACCTGTGAAGATCAAGAGGACCTGTGAGC	1183
Db	1067	AATCCGAGAGGCCACAGAGGACCTAAACCTGTGAAGATCAAGAGGACCTGTGAGC	1126
Qy	1184	GACATCACTTTTCTGTGCTGAGGAGCTGAGGCTGACCTGTCTCTGAGAGCACTCA	1243
Db	1127	GACATCACTTTTCTGTGCTGAGGAGCTGAGGCTGACCTGTCTCTGAGAGCACTCA	1186
Qy	1244	CTGCTATATGAGGAGTGTGTTGGGCTCAGAGCGAAGCTTCCCTCTCTTGAATCTGGAGTTGAA	1303
Db	1187	TTACCCATTTGGGTCTCTCGGGCTCAGAGTGAGCTTCCCTCTCTTGAATCTGGAGTTGAA	1246
Qy	1304	GCCTTACCACAGGCTTCAAGTGCAGAGGTAATGCTTCTCTTGAATCTGGAGTTGAA	1363
Db	1247	GCCTTACCACAGGCTTCAAGTGCAGAGGTAATGCTTCTCTTGAATCTGGAGTTGAA	1306
Qy	1364	GTGAAATGAGGCTTCAAGAAAGTGAAGAGCAATGCTCTCTGGGCAATGCTGCTGGC	1423
Db	1307	GTGAAATGAGGCTTCAAGAAAGTGAAGAGCAATGCTCTCTGGGCAATGCTGCTGGC	1366
Qy	1424	AGTATGCTCTTCAGAGGAGCTATGTCAGGCAATGAGTGAAGCTGGGATTCCTCAGAGCCCT	1483
Db	1367	AGTATGCTCTTCAGAGGAGCTATGTCAGGCAATGAGTGAAGCTGGGATTCCTCAGAGCCCT	1426
Qy	1484	GATGACTCAGATAGCAGCTATGTTCCCTCCTCAGGCTCAGAGCTCAGGCTGCTCCCT	1543
Db	1427	GAGGACTCAGACAGCTATGTTCCCTCCTCAGGCTCAGAGCTCAGGCTGCTCCCT	1486
Qy	1544	GTCTTCAACGAGGCTGCAAGAGAGGATGAGGAAATATAAAGGAAA	1592
Db	1487	GTCTTCAACGAGGCTGCAAGAGAGGATGAGGAAATATAAAGGAAA	1535
RESULT 2			
BM921322			
LOCUS			
DEFINITION	BM921322	1027 bp mRNA linear EST 12-MAR-2002	
ACCESSION	AGENCOURT_6626445	NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5752725	
VERSION	BM921322	5', mRNA sequence.	
KEYWORDS	BM921322.1	GI:19371701	
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: Life Technologies, Inc.		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.llnl.gov		
	Plate: L1A12787 row: f column: 22		
	High quality sequence stop: 674.		
FEATURES	Location/Qualifiers		
source	1..1027		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5752725"		
	/lab_host="NIH_MGC_115"		
	/note="Organ: pooled brain, lung, testis; Vector:		

pcmv-sport6; site_1: NotI; site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."

BASE COUNT 230 a 307 c 258 g 231 t 1 others

Query Match	55.0%	Score 940.4	DB 14	Length 1027	
Best Local Similarity	96.6%	Pred. No. 8e-162			
Matches	982	Conservative 0	Mismatches 32	Indels 3	Gaps 2
QY	2	CGGATCTCTTGGCTCAGGCGCTCTCGAGGTCACAGACAGCGCGCCAGCCGCTCTGCGACGC	61		
Db	9	CGGATCTCTTGGCTCAGGCGCTCTCGAGGTCACAGACAGCGCGCCAGCCGCTCTGCGACGC	68		
QY	62	AGCAGTGAATPAGTGTGGTACCTCTTGTCTCGGTTTCAGGTCACAGACCTCCCCGCTCTTC	121		
Db	69	AGCAGTGAATPAGTGTGGTACCTCTTGTCTCGGTTTCAGGTCACAGACCTCCCCGCTCTTC	128		
QY	122	GCTGCCCTGAACGTCAGGCGACCTCAGGACCCCTGTGATTTGGCGCCCTCGCGCGCGACCG	181		
Db	129	GCTGCCCTGAACGTCAGGCGACCTCAGGACCCCTGTGATTTGGCGCCCTCGCGCGCGACCG	188		
QY	182	TGACCGAGGAAACCCCTGGAGGACATTGGGCATTCCTTTGGGCTCCGCTGCTGTTCTTCGT	241		
Db	189	TGACCGAGGAAACCCCTGGAGGACATTGGGCATTCCTTTGGGCTCCGCTGCTGTTCTTCGT	248		
QY	242	GCTCCTTTTCGGGCAAGGATCTCACATTTATCAGTCTTTTGACCGCACAGAGATGCTCGCAT	301		
Db	249	GCTCCTTTTCGGGCAAGGATCTCACATTTATCAGTCTTTTGACCGCACAGAGATGCTCGCAT	308		
QY	302	TTGATAAATGTTTGTGAACCTTGAAGAGACATATGGACAATGAATCTGCAAGATACTGG	361		
Db	309	TTGATAAATGTTTGTGAACCTTGAAGAGACATATGGACAATGAATCTGCAAGATACTGG	368		
QY	362	GGAGAGATACCAATATCATCAAGCCAGACCAACAAGATTTCCTTCGATTTGCTCCACCG	421		
Db	369	GGAGAGATACCAATATCATCAAGCCAGACCAACAAGATTTCCTTCGATTTGCTCCACCG	428		
QY	422	GAGTTTCGCTGTGTGGAAGTCCATGATCCACCCCTGCACCAACCCCTCAGCACAACAGCG	481		
Db	429	GAGTTTCGCTGTGTGGAAGTCCATGATCCACCCCTGCACCAACCCCTCAGCACAACAGCG	488		
QY	482	AAGCCCCCTACTATGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATTCCATACGATT	541		
Db	489	AAGCCCCCTACTATGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATTCCATACGATT	548		
QY	542	CAGTTGATTCAGCACAAACCGACCTCTTCGCAACCTTATTGCCACAGCTCAGGCCCCAGAAT	601		
Db	549	CAGTTGATTCAGCACAAACCGACCTCTTCGCAACCTTATTGCCACAGCTCAGGCCCCAGAAT	608		
QY	602	CAGCAGCAGACGAAGGTGTAAAGAGTGAAGAGAGTGAACCTCTTCCCTCGTCCCTCGG	661		
Db	609	CAGCAGCAGACGAAGGTGTAAAGAGTGAAGAGAGTGAACCTCTTCCCTCGTCCCTCGG	668		
QY	662	TCACCTCCTCTCCCTGATGACCTCTCGGCTTTAGATTTGAAGAATCCCAATGACCATTC	721		
Db	669	TCACCTCCTCTCCCTGATGACCTCTCGGCTTTAGATTTGAAGAATCCCAATGACCATTC	728		
QY	722	CAGATCGGCGACAGTACCAGAGAGTGACTTTTATGCTGGGAAAGGGAACCTGTGACT	781		
Db	729	CAGATCGGCGACAGTACCAGAGAGTGACTTTTATGCTGGGAAAGGGAACCTGTGACT	788		
QY	782	GAACTAGCTGGGCACTCTCTCGGAGCTCTCTTACAGGAGCTGCGCCACAATCTCGGCC	841		
Db	789	GAACTAGCTGGGCACTCTCTCGGAGCTCTCTTACAGGAGCTGCGCCACAATCTCGGCC	848		
QY	842	CACGCGGGCTTTGACTGTGCTAATGAGAGTGTCTCTGGAGACCCCTAACTGATGTGGCACAT	901		

```

849 CACCGGGCTTTGACTGCTAATGAGAATGCCTGGAGACCT-ACTGATGTGCACCT 907
|||||
902 GAGTATTCCTTAAGTTACCAAGTTGCTGCTTTTGGTGTGGACCGGAGCCCGGCTG 961
|||||
908 GAATATGCCCTTAA--TTTCAAGTGGCTCGTTTCTTGGGACCGGAAGGCCGCTT 965
|||||
962 GGACAGACTCCTTTTCTCTGATGATGAGCAGGATTTCCATGAAGTGGGTATGGC 1018
|||||
966 GTGAAACTCCTTTTCTCTGAAGGAAGGACAGGATTCCTTTGAAGGGGTATTGCC 1022
|||||

RESULT 3
BM545491 939 bp mRNA linear EST 20-FEB-2002
LOCUS AGENCOURT_6500700 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5729706
5', mRNA sequence.
ACCESSION BM545491
VERSION BM545491.1 GI:18777588
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 939)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12727 row: g column: 19
High quality sequence stop: 634.

FEATURES
source
1. 939
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5729706"
/clone_lib="NIH_MGC_124"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: EcorV
(destroyed); Site_2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcorV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012."

BASE COUNT 211 a 283 c 234 g 209 t 2 others
ORIGIN
Query Match 52.9%; Score 904.6; DB 13; Length 939;
Best Local Similarity 97.8%; Pred. No. 2.8e-155;
Matches 916; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 27 AGGTCCAGACACGCCGCCCTCTGCCACGACGAGTGAATAGTGTGGTACCTCCT 86
DB 1 AGGTCCAGACACGCCGCCCTCTGCCACGACGAGTGAATAGTGTGGTACCTCCT 60
QY 87 TGTCTCGGTTGAGTCCAGACCTCCCGTCTTCCGGTGCCTGAAGCTCAGGCGACCTC 146
DB 61 TGTCTCGGTTGAGTCCAGACCTCCCGTCTTCCGGTGCCTGAAGCTCAGGCGACCTC 120
QY 147 AGGACCTGTGATTGGGCGCTTCGCCGCGGACCTGACCGAGGAACCCCTGCGAGGAC 206
DB 121 AGGACCTGTGATTGGGCGCTTCGCCGCGGACCTGACCGAGGAACCCCTGCGAGGAC 180

```

```

QY 207 TTGGCATTCCTTGGGCTCCGTCCTGCTGCTTCTTCTGCTCCTTTGGGCAAGGATCTCACA 266
|||||
DB 181 TTGGCATTCCTTGGGCTCCGTCCTGCTGCTTCTTCTGCTCCTTTGGGCAAGGATCTCACA 240
|||||
QY 267 TTATCAGCTTTTGGACCGACACAGAAATGCTCGGATTTGATTAATGTTGTTGAACCTGAA 326
|||||
DB 241 TTATCAGCTTTTGGACCGACACAGAAATGCTCGGATTTGATTAATGTTGTTGAACCTGAA 300
|||||
QY 327 GAGACATATGACAAATGAATCTGCAAGATACTGGGAGAGATACCAATATATCATCAAGCC 386
|||||
DB 301 GAGACATATGACAAATGAATCTGCAAGATACTGGGAGAGATACCAATATATCATCAAGCC 360
|||||
QY 387 AGACCAACAGAAAGTTCTTCTGATTTGCTCCACGCGGAGTTCCTGCTGTTGGAAGTCCATG 446
|||||
DB 361 AGACCAACAGAAAGTTCTTCTGATTTGCTCCACGCGGAGTTCCTGCTGTTGGAAGTCCATG 420
|||||
QY 447 ACCCACCCCTGCACCAACCCCTCAGCCCAACAGCGGAGCCGCCCATATGCTGGACATCC 506
|||||
DB 421 ACCCACCCCTGCACCAACCCCTCAGCCCAACAGCGGAGCCGCCCATATGCTGGACATCC 480
|||||
QY 507 CCTCAGAGCCATGATGCTCACCATCCATAGATTGATTCAGTTTCCAGACACACGACCTC 566
|||||
DB 481 CCTCAGAGCCATGATGCTCACCATCCATAGATTGATTCAGTTTCCAGACACACGACCTC 540
|||||
QY 567 TTCGCAACCTTATTGCCACAGCTCAGGCCCAAGATTCAGCAGCAGACAGAGAGTGTAATAA 626
|||||
DB 541 TTCGCAACCTTATTGCCACAGCTCAGGCCCAAGATTCAGCAGCAGACAGAGAGTGTAATAA 600
|||||
QY 627 CTGAAGAGAGTAACCTTCTCCCTCGTGGCTCGGTCACCTCCTCTCTCTGATGACCTC 686
|||||
DB 601 CTGAAGAGAGTAACCTTCTCCCTCGTGGCTCGGTCACCTCCTCTCTCTGATGACCTC 660
|||||
QY 687 TGCCTTTAGATTGTAAGAAATCCCAATGCCATTCAGATCCGACAGTCCGACAGAGAGA 746
|||||
DB 661 TGCCTTTAGATTGTAAGAAATCCCAATGCCATTCAGATCCGACAGTCCGACAGAGAGA 720
|||||
QY 747 GTGACTTTTATCGTGGGAAAGGGAACCTGTGACTGAACTCAGCTGGCCTCTCTCTGCTGCGC 806
|||||
DB 721 GTGACTTTTATCGTGGGAAAGGGAACCTGTGACTGAACTCAGCTGGCCTCTCTCTGCTGCGC 780
|||||
QY 807 AGCTCCTCTACGAGCAGTGGCCACATCTTGGCCCGGAGGCTTTGACTGTGCTAATG 866
|||||
DB 781 AGCTCCTCTACGAGCAGTGGCCACATCTTGGCCCGGAGGCTTTGACTGTGCTAATG 840
|||||
QY 867 AGAGTGTCTCGAGACCTTAACCTGATGTGGCAGATGATTTGCTTAAGTTTACCAAGT 926
|||||
DB 841 AGAGTGTCTCGAGACCTTAACCTGATGTGGCAGATGATTTGCTTAAGTTTACCAAGT 900
|||||
QY 927 TGCTGCGTTTGTGTGAGCCGGGAGGCCCGGCTGGG 963.
|||||
DB 901 TGCTGCGTTTGTGTGAGCCGGGAGGCCCGGCTGGG 937
|||||

```

RESULT 4

BQ229669

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BQ229669

AGENCOURT_7553581 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6046058

5', mRNA sequence.

BQ229669

BQ229669.1 GI:20411069

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 865)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LLM13291 row: e column: 03
High quality sequence stop: 653.
Location/Qualifiers
1. .865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6046058"
/clone_lib="NIH_MGC_92"
/tissue_types="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: Sall; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
214 a 243 c 210 g 197 t 1 others

BASE COUNT 214 a 243 c 210 g 197 t 1 others
ORIGIN

Query Match 48.2%; Score 824.4; DB 14; Length 865;
Best Local Similarity 99.4%; Pred. No. 1.2e-140;
Matches 858; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 370 ACCAATATCATCAAGCCAGACACAGAGTCTCTCGATTGCTCCACGGGAGTTCCG 429
|||||
Db 1 ACCAATATCATCAAGCCAGACACAGAGTCTCTCGATTGCTCCACGGGAGTTCCG 60
|||||

QY 430 TCTGTGGAGTCCATGACCCACCCCTCGACCAACCTCTAGCCAAAGCCGAAGCCCC 489
|||||
Db 61 TCTGTGGAGTCCATGACCCACCCCTCGACCAACCTCTAGCCAAAGCCGAAGCCCC 120
|||||

QY 490 CACTATGCTGGACATCCCTCAGAGCATGTAGTCTCACCATCCATACAGTTGAT 549
|||||
Db 121 CACTATGCTGGACATCCCTCAGAGCATGTAGTCTCACCATCCATACAGTTGAT 180
|||||

QY 550 TCAGCACACCGAGCTCTTGGCAACCTTATGTCACAGCTCAGGCCACCAATCAGCAGCA 609
|||||
Db 181 TCAGCACACCGAGCTCTTGGCAACCTTATGTCACAGCTCAGGCCACCAATCAGCAGCA 240
|||||

QY 610 GACAGAAGGTGTAAACTGAAGAGAGTGAACCTCTTCCCTCGTCCCTGGTCACTCC 669
|||||
Db 241 GACAGAAGGTGTAAACTGAAGAGAGTGAACCTCTTCCCTCGTCCCTGGTCACTCC 300
|||||

QY 670 TCTCCCTGATGACCTCTGCTGTTAGATTGTAAAGATCCCAATGCCACCAATCCAGATCCG 729
|||||
Db 301 TCTCCCTGATGACCTCTGCTGTTAGATTGTAAAGATCCCAATGCCACCAATCCAGATCCG 360
|||||

QY 730 GCACACTGACCCAGAGAGTCACTTTATCTGTGGAAAGGGAGACCTGTGACTGAATCAG 789
|||||
Db 361 GCACACTGACCCAGAGAGTCACTTTATCTGTGGAAAGGGAGACCTGTGACTGAATCAG 420
|||||

QY 790 CTGGCACTCTGTGCGCAGCTCTCTACAGGAGTGGCCACCAATCTGCCCCACCGGG 849
|||||
Db 421 CTGGCACTCTGTGCGCAGCTCTCTACAGGAGTGGCCACCAATCTGCCCCACCGGG 480
|||||

QY 850 CTTTGACTGTGTAATAGAGTGTCTGGAGACCCCTAACTGATGTGGCAGATGATTTG 909
|||||
Db 481 CTTTGACTGTGTAATAGAGTGTCTGGAGACCCCTAACTGATGTGGCAGATGATTTG 540
|||||

QY 910 CCTTAAGTTTACCAAGTTGCTGCTGTTGCTGTGGACCGGGAGCCGGCTGGACAGAC 969
|||||
Db 541 CCTTAAGTTTACCAAGTTGCTGCTGTTGCTGTGGACCGGGAGCCGGCTGGACAGAC 600
|||||

QY 970 TCTTTTCTGTGATGTGATGAGCAGGTATTCCATGAAGTGGGTATTGGCAGTGTGCTTC 1029
|||||
Db 601 TCTTTTCTGTGATGTGATGAGCAGGTATTCCATGAAGTGGGTATTGGCAGTGTGCTTC 660
|||||

QY 1030 CTTCCAGAGTTCTTGGCAGACCGGCATCAAGGACTATCAGATTACATGCTACAGATTAG 1089
|||||

|||||
Db 661 CTCCCAAGTCTTGGCAGACCCGATCAAGGACTATCAGGTATACAGTATAGATTAG 720
|||||

QY 1090 TAAGCAACTCTCTGAAGAATATGAAGGATTGTCAATCCCTGAGAAGCCACAGAGGACG 1149
|||||

Db 721 TAAGCAACTCTCTGGAGAATATGAAGGATTGTCAATCCCTGAGAAGCCACAGAGGACG 780
|||||

QY 1150 TAAACCTGTGAAGATCAA-GGAGGAACCTCTGAGCGACATCAC-TTTTCTGTGAGT-GA 1206
|||||

Db 781 TAAACCTGTGAAGATCAAAGGGAGGAGACCTGTGAGCGACATCACTTTTCTGTGTCANTGA 840
|||||

QY 1207 GGAGCTGGAGGCTGACCTTCTT 1229
|||||

Db 841 GGAGCTGGAGGCTGACCTTCTT 863
|||||

RESULT 5
BM545988 1214 bp mRNA linear EST 20-FEB-2002
LOCUS
DEFINITION AGENCOURT_6505156 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5588387
5', mRNA sequence.
ACCESSION BM545988
VERSION BM545988.1 GI:18778575
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1214)
NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LLM12359 row: g column: 12
High quality sequence start: 135
High quality sequence stop: 529.
Location/Qualifiers
1. .1214
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5588387"
/clone_lib="NIH_MGC_125"
/lab_host="DH10B"
/notes="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool
of three ovaries, from females ranging in age from 38 to
49 yo. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 2.1 kb, insert size range 1-3.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 036."
268 a 347 g 312 g 250 t 37 others

BASE COUNT 268 a 347 g 312 g 250 t 37 others
ORIGIN

Query Match 47.8%; Score 817.4; DB 13; Length 1214;
Best Local Similarity 97.4%; Pred. No. 2e-139;
Matches 827; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 ACAGCATCTTGGCTCAGGCTCTCGAGGTCCAGACAGCCGCCAGCCGCTCTGCGACG 60
|||||

Db 116 ACCCNCANCTTCCCTCANGCTCTCNAGGTCCAAACAGCCGCCAGCCGCTCTGCGACG 175
|||||

QY 61 CAGCAGTAATAGTGTGGTACCTCTTGTCTCGTTCAGGTCCAGACCTCCCGCTTCC 120
|||||

```
Db 176 CNGCAGTGAATAGTGTGGTACCTCTCTGTCTCGGTTCAAGTCCAGACCTCCCGCTTCC 235
QY 121 GGCTGCTTGAACGTCAGGCGACCTCAGGACCTCTGTATTGCGCTGCGCGCGGAC 180
Db 236 GGCTGCTTGAACGTCAGGCGACCTCAGGACCTCTGTATTGCGCTGCGCGCGGAC 295
QY 181 GTGACCGAGGAACCCCTGGAGGACTTGGGCAATCTTGGGCTCCGCTGCTTCTTCG 240
Db 296 GTGACCGAGGAACCCCTGGAGGACTTGGGCAATCTTGGGCTCCGCTGCTTCTTCG 355
QY 241 TCGTCTCTTCGGCGAAGNCTCAGATTATCAGTCTTTGACCGACAGCAAGTCCGCGCA 300
Db 356 TCGTCTCTTCGGCGAAGNCTCAGATTATCAGTCTTTGACCGACAGCAAGTCCGCGCA 415
QY 301 TTTGATAAATGTTTGTGAACCTTGAAGACATATGGCAATGAATCTGCAAGATACGT 360
Db 416 TTTGATAAATGTTTGTGAACCTTGAAGACATATGGCAATGAATCTGCAAGATACGT 475
QY 361 GGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCTTTCGATTGCTCCACG 420
Db 476 GGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCTTTCGATTGCTCCACG 535
QY 421 GGAGTTCCGCTGCTGGTGGAGTCCATGACCCCTGACCAACCCCTCAGCCCAAGGCC 480
Db 536 GGAGTTCCGCTGCTGGTGGAGTCCATGACCCCTGACCAACCCCTCAGCCCAAGGCC 595
QY 481 GAAGCCCCCCTACTGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCCATACGAT 540
Db 596 GAAGCCCCCCTACTGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCCATACGAT 655
QY 541 TCAGTTGATTCAGCACACACCGACGCTTTCGCAACCTTATGCCACAGCTCAGGCCAGAA 600
Db 656 TCAGTTGATTCAGCACACACCGACGCTTTCGCAACCTTATGCCACAGCTCAGGCCAGAA 715
QY 601 TCAGCAGCAGACAGAAAGCTGTAAACCTGAGAGAGTGAACCTCTTCCCTGCGCCCTGG 660
Db 716 TCAGCAGCAGACAGAAAGCTGTAAACCTGAGAGAGTGAACCTCTTCCCTGCGCCCTGG 775
QY 661 CTCACCTCTCTCCCTGATGACCTCTGCTCTTATGATTGTAAGATCCCAATGACACCAT 720
Db 776 CTCACCTCTCTCCCTGATGACCTCTGCTCTTATGATTGTAAGATCCCAATGACACCAT 835
QY 721 CCAGATCCGCGACAGTACCCAGAGAGTACTTTTATCGTGGGAAAGGGACCTGTGAC 780
Db 836 CCAGATCCGCGACAGTACCCAGAGAGTACTTTTATCGTGGGAAAGGGACCTGTGAC 895
QY 781 TGAACCTCAGCTGGCACTCTCTGCGCAGCTCTCTTACCAGCAGTGGCCACAACTCCTGC 840
Db 896 TGAACCTCAGCTGGCACTCTCTGCGCAGCTCTCTTACCAGCAGTGGCCACAACTCCTGC 955
QY 841 CCACCGGG 849
Db 956 CCCACCGG 964

RESULT 6
BQ924589
LOCUS
DEFINITION BQ924589 923 bp mRNA linear EST 20-AUG-2002
5', mRNA sequence.
ACCESSION BQ924589
VERSION BQ924589.1 GI:22339620
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 923)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
```

```
Tissue Procurement: DCTD/DTP/Gazdar
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2600 row: c column: 19
High quality sequence stop: 635.
location/Qualifiers
1. 923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 209 a 283 c 221 g 210 t
ORIGIN
Query Match 46.6%; Score 797.4; DB 14; Length 923;
Best Local Similarity 96.7%; Pred. No. 9.6e-136;
Matches 889; Conservative 0; Mismatches 21; Indels 9; Gaps 7;
QY 27 AGGTCCACACAGCCGCCCGCTCTGCGACGACGAGTGAATAGTGTGACTCTCT 86
Db 1 AGTCCACACAGCCGCCCGCTCTGCGACGACGAGTGAATAGTGTGACTCTCT 60
QY 87 TGTCTCGTTCAGGTCCAGACCTCCCGCTCTTCCGGCTGCCCTGAACTGACGGCACCTC 146
Db 61 TGTCTCGTTCAGGTCCAGACCTCCCGCTCTTCCGGCTGCCCTGAACTGACGGCACCTC 120
QY 147 AGGACCTGTATTGGCGCTTGGCGCGGACCGTACCAGGAGAAACCCCTGGAGGAC 206
Db 121 AGGACCTGTATTGGCGCTTGGCGCGGACCGTACCAGGAGAAACCCCTGGAGGAC 180
QY 207 TTGGCATTCTTGGGCTCCGCTGCTTCTTCTGCTGCTTCTGCGGCAAGGATCTCACA 266
Db 181 TTGGCATTCTTGGGCTCCGCTGCTTCTTCTGCTGCTTCTTCCGGCAAGGATCTCACA 240
QY 267 TTATCAGTCTTTGACCGACACAGAAATGCTGGCAATTTGATAAATGTTTGTGAACCTGAA 326
Db 241 TTATCAGTCTTTGACCGACACAGAAATGCTGGCAATTTGATAAATGTTTGTGAACCTGAA 300
QY 327 GAGACATATGACATGATGATCTGCAAGATACTGGGAGAGATACCAATATCATCAAGCC 386
Db 301 GAGACATATGACATGATGATCTGCAAGATACTGGGAGAGATACCAATATCATCAAGCC 360
QY 387 AGACCAACAGAAAGTCTTCTGATTTGCTCCACGCGGAGTTCGCTGCTGCTGGAAGTCCATG 446
Db 361 AGACCAACAGAAAGTCTTCTGATTTGCTCCACGCGGAGTTCGCTGCTGGAAGTCCATG 420
QY 447 ACCACCCCTGCACCAACCCCTCAGCCCAACAGCCGAGCCGCCCACTACTGTGGACATCC 506
Db 421 ACCACCCCTGCACCAACCCCTCAGCCCAACAGCCGAGCCGCCCACTACTGTGGACATCC 480
QY 507 CCTCAGAGCCATGTAGTCTCACCATCCATAGATTTCAGTTCAGTTCAGACACACCGACCTC 566
Db 481 CCTCAGAGCCATGTAGTCTCACCATCCATAGATTTCAGTTCAGTTCAGACACACCGACCTC 540
QY 567 TTCGCAACCTTATTGGCCACAGCTCAGGCCCGGAGAGTTCAGCAGCAGACAGAAAGTGTAAAA 626
Db 541 TTCGCAACCTTATTGGCCACAGCTCAGGCCCGGAGAGTTCAGCAGCAGACAGAAAGTGTAAAA 600
QY 627 CTGAAGAGAGTGAACCTCTTCCCTGCTGCGCTGGGTACCTCTCTCCCTGATGA-CCTC 685
```

		Matches	858;	Conservative	0;	Mismatches	36;	Indels	8;	Gaps	7;
Qy	215	TCCTTGGCGTCCGTCGCTGTTCTTCGTCGCTCCTTTTCGGGCAAGGATCTCACATTATCAGT	274								
Db	3	TTCTCTGGGCTCCGTCGCTGTTCTTCGTCGCTCCTTTTCGGGCAAGGATCTCACATTATCAGT	62								
Qy	275	CTTTGACCGACACAGAAATGCTGGCATTTTGTAAATGTTTGTGAATCTTGAAGAGACATA	334								
Db	63	CTTTGACCGACACAGAAATGCTGGCATTTTGTAAATGTTTGTGAATCTTGAAGAGACATA	122								
Qy	335	TGGACAATGAATCTGCAAAAGATACTGGGGAGAGATACCAATATCATCAAGCCAGACCAAC	394								
Db	123	TGGACAATGAATCTGCAAAAGATACTGGGGAGAGATACCAATATCATCAAGCCAGACCAAC	182								
Qy	395	AGAAGTTCCTTTGCTTTGCTCCACGGGAGTTCGCTGCTGGTGGGAAGTCCATGACCCACCC	454								
Db	183	AGAAGTTCCTTTGCTTTGCTCCACGGGAGTTCGCTGCTGGTGGGAAGTCCATGACCCACCC	242								
Qy	455	CTGCACCAACCCCTCAGCCAAACAGCCGGAAGCCCCCCTATGCTGGACATCCCTCCAGAG	514								
Db	243	CTGCACCAACCCCTCAGCCAAACAGCCGGAAGCCCCCCTATGCTGGACATCCCTCCAGAG	302								
Qy	515	CCATGTAGTCTCACCATCCATAGATTTCAGTTTTCAGTTCAGACACACCCAGCTCTTCGCAAC	574								
Db	303	CCATGTAGTCTCACCATCCATAGATTTCAGTTTTCAGTTCAGACACACCCAGCTCTTCGCAAC	362								
Qy	575	CTTATTTCCACAGCTCAGGCCCAAGAAATCAGCAGCAGACAGAAAGTGTAATAAAGTGAAGAG	634								
Db	363	CTTATTTCCACAGCTCAGGCCCAAGAAATCAGCAGCAGACAGAAAGTGTAATAAAGTGAAGAG	422								
Qy	635	AGTGAACCTCTTCCCTCGTGGCTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT	694								
Db	423	AGTGAACCTCTTCCCTCGTGGCTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT	482								
Qy	695	GATTGTAAGAAATCCCAATGCACCAATCCAGATCCGGCAGACAGTACCACAGAGAGTGACTTT	754								
Db	483	GATTGTAAGAAATCCCAATGCACCAATCCAGATCCGGCAGACAGTACCACAGAGAGTGACTTT	542								
Qy	755	TATCTGTTGGGAAAGGGGAACTGTGACT-GAACTCAGCTGGCAGTCTCCTGTGCGGAGTCTCT	813								
Db	543	TATCTGTTGGGAAAGGGGAACTGTGACT-GAACTCAGCTGGCAGTCTCCTGTGCGGAGTCTCT	602								
Qy	814	CTACCAAGCAGTGGCCACAA--TCCTGGCCCCACCGCGGCTTTCAC-TGTGCTAATGAGAG	870								
Db	603	CTACCAAGCAGTGGCCACAAATCCTGGCCCCACCGCGGCTTTCAC-TGTGCTAATGAGAG	662								
Qy	871	TGTCCTGGAGACCCCTTACTGATCTGSCACATGAGTATTGCTTTAAGTTTACCAAGTTGCT	930								
Db	663	TGTCCTGGAGA-CCTAACTGATGTGCACCATGAGTATGCTTTAGTTTACCAAGTTGCT	721								
Qy	931	GCCTTTTGTGTCGACCGCGGAGCGCGCTGGGA-CAGAGTCCCTTTTCCCTGATGTGATGG	989								
Db	722	GCCTTTTGTGTCGACCGCGGAGCGCGCTGGGA-CAGAGTCCCTTTTCCCTGATGTGATGG	781								
Qy	990	-AGCAGGTAATTCATGAAGTGGGTATTTGGCAGTGTGCTCTCCCTCCAGAAAGTTCTGGCAG	1048								
Db	782	CAGCAGGTAATTCATGAAGTGGGTATTTGGCAGTGTGCTCTCCCTCCAGAAAGTTCTGGCAG	841								
Qy	1049	CACCGCATCAAGCACTATCACAGTTACATGCTACA-GATTAGTAAGCAACTCTCTGAAGA	1107								
Db	842	CACCGGATTCAGGACTATAACAGTCACTGGCTACAGGAATAGTAAGCAACTCTCTGAAGA	901								
Qy	1108	AT 1109									
Db	902	AT 903									
RESULT 8											
LOCUS	BI669780	818 bp mRNA linear EST 12-SEP-2001									
DEFINITION	603293336F1 NIH_MGC_96	Homo sapiens cDNA clone IMAGE:5312869 5'									
ACCESION	BI669780	mRNA sequence.									

Db	601	CTGAAGAGAGTGAACCTTTCCCTCGCGCTCCTCCCTGATGACCCTC	660								
Qy	686	CTGCCTTTTAGATTGAAGAATCCCAATGACCATTCAGATCCGGCACAGTGACCCAGAG	745								
Db	661	CTGCCTTTTAGATTGTAGAATGCCAATGACCATTCAGATCCGGCACAGTGACCCAGAG	720								
Qy	746	AGTGACTTTT--ATCGTGGGAAAGGGAACCTGTGACTGAACTGGCAGCTCCTGTC	803								
Db	721	AGTGACTTTTTATCGCTGGGAAAGGGAACCTGTGACTGAACTCACCTGGAACCTCCTGTT	780								
Qy	804	GGCAGCTCCTTACAGGCAGTGCCACAATCCTGGCCCCAGCG--GGGCTTTGACTGTG-C	861								
Db	781	GGCAGCTCCTCTACC--GGGAGTGCCCAATCCTGGCCCCAGCGGGGCTTTGACTGTGCC	839								
Qy	862	TAATGAGAGTGTCTCT-GGAGACCCCTAACTGATGTGGCACATGAGTATTGGC--TTAAGTT	918								
Db	840	TAATGAAGTGGCTGGGAACCCCTAACTGATGTGGCCCCTGAATATTGCCCTTAATTT	899								
Qy	919	TACCAAGTGTGCGTTTT 937									
Db	900	TCCAAGTTGCGCTGCTTT 918									
RESULT 7											
LOCUS	BI822539	904 bp mRNA linear EST 04-OCT-2001									
DEFINITION	603035248F1 NIH_MGC_115	Homo sapiens cDNA clone IMAGE:5176547 5'									
ACCESION	BI822539	mRNA sequence.									
VERSION	BI822539.1	GI:15934089									
KEYWORDS	EST.										
SOURCE	human.										
ORGANISM	Homo sapiens										
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;											
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.											
1 (bases 1 to 904)											
NIH-MGC http://mgs.nci.nih.gov/.											
National Institutes of Health, Mammalian Gene Collection (MGC)											
Unpublished (1999)											
Contact: Robert Strausberg, Ph.D.											
Email: ccapbs-re@mail.nih.gov											
Tissue Procurement: Life Technologies, Inc.											
cDNA Library Preparation: Life Technologies, Inc.											
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)											
DNA Sequencing by: Incyte Genomics, Inc.											
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:											
http://image.llnl.gov											
Plate: LLAM1439 row: o column: 12											
High quality sequence stop: 825.											
Location/Qualifiers											
1..904											
/organism="Homo sapiens"											
/db_xref="taxon:9606"											
/clone="IMAGE:5176547"											
/clone.lib="NIH_MGC_115"											
/lab_host="DH10B"											
/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 Kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."											
BASE COUNT	219 a	252 c	220 g	213 t							
ORIGIN											
Query Match	44.3%;	Score	758.4;	DB 13;	Length	904;					
Best Local Similarity	95.1%;	Pred.	No. 1.2e-128;								

```
VERSION BI669780.1 GI:15584013
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 818)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11792 row: o column: 14
High quality sequence stop: 754.
FEATURES
source
1..818
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcggag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 196 a 234 c 204 g 184 t
ORIGIN
Query Match 43.9%; Score 750.6; DB 13; Length 818;
Best Local Similarity 98.3%; Pred. No. 3.4e-127;
Matches 801; Conservative 0; Mismatches 9; Indels 5; Gaps 4;
QY 161 GGCCTCGCCGCGGCGGACCTGACCGAGGAACCCCTGGAGGACTTGGCATTCCTTG 220
Db 6 GCGCCTCGCGCGGCGGACCTGACCGAGGAACCCCTGGAGGACTTGGCATTCCTTG 65
QY 221 GGCCTCGCCGCTGCTTCGCTGCTTCCTTCGGCAAGGATCTACATATCAGTCTTGA 280
Db 66 GGCCTCGCCGCTGCTTCGCTGCTTCCTTCGGCAAGGATCTACATATCAGTCTTGA 125
QY 281 CCGACACAGATGCTGCGCATTTGATAAATGTTTGTGAAGTTCGAAGACATATGGACA 340
Db 126 CCGACACAGATGCTGCGCATTTGATAAATGTTTGTGAAGTTCGAAGACATATGGACA 185
QY 341 ATGAATCTGCAAGATACCTGGGAGAGATACCATATATCATCAAGCCAGACAGAGT 400
Db 186 ATGAATCTGCAAGATACCTGGGAGAGATACCATATATCATCAAGCCAGACAGAGT 245
QY 401 TCCCTTCGATTTGCTCCACGGGAGTTCGCTGCTGCTGGAAGTCCATGACCCCTGCAC 460
Db 246 TCCCTTCGATTTGCTCCACGGGAGTTCGCTGCTGCTGGAAGTCCATGACCCCTGCAC 305
QY 461 CAACCCCTCAGCAACAAGCCGGAAGCCGCCACTATGTTGGACATCCCTCAGAGCCATGT 520
Db 306 CAACCCCTCAGCAACAAGCCGGAAGCCGCCACTATGTTGGACATCCCTCAGAGCCATGT 365
QY 521 AGTCTCACCATTCCATGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT 580
Db 366 AGTCTCACCATTCCATGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT 425
```

```
581 GCCACAGCTCAGGCCGCCAGAAATCAGCAGCAGACAGAGAGTGTTAAAACTGAAGAGAGTGA 640
|||||
426 GCCACAGCTCAGGCCGCCAGAAATCAGCAGCAGACAGAGAGTGTTAAAACTGAAGAGAGTGA 485
|||||
641 CCTCTTCCCTCGTCCCTGGGTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 700
|||||
486 CCTCTTCCCTCGTCCCTGGGTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 545
|||||
701 AAGAAATCCCAATGACCAATCCAGATCCGACAGTCCGACAGTCCGACAGAGAGTGTATTCGT 760
|||||
546 AAGAAATCCCAATGACCAATCCAGATCCGACAGTCCGACAGTCCGACAGAGAGTGTATTCGT 605
|||||
761 GGGAAAGGGGAACTGTGACTG-AACTCAGTGGCAGTCTCTGTCGGCAGCTCCTCTACCA 819
|||||
606 GGGAAAGGGGAACTGTGACTGAAACTCAGTGGCAGTCTCTGTCGGCAGCTCCTCTACCA 665
|||||
820 GGCAGTGGCCAC-AATCCTGGCCACAGGGGCTTTGACTG-TGCTAATGAGAGTGTCTCTG 877
|||||
666 GGCAGTGGCCACAAATCCTGGCCACAGGGGCTTTGACTG-TGCTAATGAGAGTGTCTCTG 725
|||||
878 GAGACCTTAAGTGTGGCAGATGATTTGCTTAAAGTTTACCAAGTTCCTGCTGCTTTT 937
|||||
726 GAGACCTTAAGTGTGGCAGATGATTTGCTTAAAGTTTACCAAGTTCCTGCTGCTTTT 783
|||||
938 GCTGTGGACCGGAGGCGCGCTGGGACAGACTCC 972
|||||
784 GCTGTGGACCGGAGGCGCGCTGGGACAGACTTC 818
|||||
RESULT 9
BM468753
LOCUS BM468753
DEFINITION AGENCOURT_6481158 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5541435
5', mRNA sequence.
ACCESSION BM468753
VERSION BM468753.1 GI:18517795
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 972)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12239 row: c column: 04
High quality sequence stop: 648.
FEATURES
source
1..972
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 251 a 265 c 237 g 217 t 2 others
ORIGIN
```


QY	1016	GGCAGTGTCTCTCCCTCCAGAAAGTTCTGGCAGCACCGCATCAAGGACTATCAAGTTAC	1075
Db	659	-GGCAGGTGCTCTCCCTCCAGAAAGTTCTGGCAGCACCGCATCAAGGACTATCAAGTTAC	717
QY	1076	ATGCTACAGATTAGTAAGCAACTCTCTGAAGA--ATATGAAGAGATTGTCAATCTCG--	1130
Db	718	ATGCTACAGATTAGTAAGCAACTCTCTGAAGACATATGGACAGGGATGGTCAATCTCTGGA	777
QY	1131	AGAAGGCCACAGAGGACGCTAA-CCTGTGAAGATCAAGGAGGAACCTGTG-AGCGACAT	1188
Db	778	GGAAAGGCCACAGAGGACGCTAAACCTGTGAAGATCAAGGAGGAACCTGTGAAGCGACAT	837
QY	1189	CACCTTTCTCTGTCAGTG-AGGAGCTGGAGGC-TGACCTTGTCTCTGGAGACCAGTCACTG	1246
Db	838	CAGCTTACCGTCAGGAAGGAGCTGGAGGATGACCTGCACCTCGGAGACCAGTGCAC	897
QY	1247	CTATGGGAGTGCTTTGGGCG 1266	
Db	898	GCTAGAGAGACGAGGGGC 917	
RESULT 11			
LOCUS	BI770385	771 bp mRNA linear EST 25-SEP-2001	
DEFINITION	603055622F1 NIH_MGC_122	Homo sapiens cDNA clone IMAGE:5205245 5', mRNA sequence.	
ACCESSION	BI770385		
VERSION	BI770385.1	GI:15761963	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM11514 row: k column: 06 High quality sequence stop: 768.		
FEATURES	Location/Qualifiers		
..	1..771		
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5205245" /clone_lib="NIH_MGC_122" /lab_host="DH10B" /note="organ: pooled lung and spleen; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library."		
BASE COUNT	175 a 243 c 184 g 168 t		
ORIGIN			
Query Match	42.2%;	Score 721.4;	DB 13; Length 771;
Best Local Similarity	98.9%;	Pred. No. 7.3e-122;	
Matches 737; Conservative	0; Mismatches	6; Indels	2; Gaps 1;

Orthopedics
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES

source

1..735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-H-EZ1-bb1-i-21-0-UI"
/clone_lib="NCI_CGAP_Ch2"
/tissue_type="Chondrosarcoma Grade II"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)
with a modified polylinker; Site.1: EcoR I; Site.2: Not I;
NCI_CGAP_Ch2 is a normalized cDNA library containing the
following tissue(s): Chondrosarcoma Grade II. The library
was constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(df)18 tail. The sequence tag for this library is
TGATCAGCT.

TAG LIB=UI-H-EZ1

TAG_TISSUE=grade-2-chondrosarcoma

TAG_SEQ=ATCTAATAG

177 a 187 c 158 g 211 t 2 others

BASE COUNT 177 a 187 c 158 g 211 t 2 others

ORIGIN

Query Match 41.5%; Score 710.8; DB 14; Length 735;

Best Local Similarity 98.6%; Pred. No. 6.4e-120;

Matches 726; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

964 ACAGACTCTTTCTGATGTGATGGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

965 ACATACTCTCTTTCTGATGTGATGGAGGAGGATTCATGAAGTGGTATGGCAGTGT 677

1024 GCTCTCCCTCCAGAGTTCTGGCAGCAGCCCATCAAGGACTATCAGATTACATGCTACA 1083

676 GCTCTCCCTCCAGAGTTCTGGCAGCAGCCCATCAAGGACTATCAGATTACATGCTACA 617

1084 GATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTTCAATCTCTGAGAGGCCACAGA 1143

616 GATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTTCAATCTCTGAGAGGCCACAGA 557

1144 GGACGCTAAACCTGTGAAGATCAAGGAGGAGCACTGTGAGCGACATCACTTTTCTGTGAC 1203

556 GGACGCTAAACCTGTGAAGATCAAGGAGGAGCACTGTGAGCGACATCACTTTTCTGTGAC 497

1204 TGAGGAGCTGGAGGCTGACCTTCTCTGGAGCAGTCACTGCTCTGAGGAGTGGTGG 1263

496 TGAGGAGCTGGAGGCTGACCTTCTCTGGAGCAGTCACTGCTCTGAGGAGTGGTGG 437

1264 GGCTCAGAGCGAAGCGTTCCATCTAACCTGGAGGTTGAAGCTTCAACACAGGCTTCAAG 1323

436 GGCTCAGAGCGAAGCGTTCCATCTAACCTGGAGGTTGAAGCTTCAACACAGGCTTCAAG 377

1324 TGACAGGTAATAGTCTCTCTCTCTTGGAAATCTGGCCCATGTGAAATGGAGCCTCAAGA 1383

376 TGACAGGTAATAGTCTCTCTCTCTTGGAAATCTGGCCCATGTGAAATGGAGCCTCAAGA 317

1384 AAGTGAAGAGGCAATGTCTCTGGGATGTGTGCTGGGAGTGTGTCTTCGAGGAGCC 1443

316 AAGTGAAGAGGCAATGTCTCTGGGATGTGTGCTGGGAGTGTGTCTTCGAGGAGCC 257

1444 TATGTCAGGATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

256 TATGTCAGGATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 197

1504 TGGTTCCACTCCACTGACAGCCTCATGGGGTCTCCCTGTTTTCACACGCGCTGCAA 1563

196 TGGTTCCACTCCACTGACAGCCTCATGGGGTCTCCCTGTTTTCACACGCGCTGCAA 137

1564 GAAGAGGATGAGGAAATATATAAGGAAAGAGGAGATGTTTGTCCAGACCTACTAGA 1623

136 GAAGAGGATGAGGAAATATATAAGGAAAGAGGAGATGTTTGTCCAGACCTACTAGA 77

1624 CCCAACAGAAAGGTTTGTATTAGAACTGTTTCTTAAATGATTGACTCTCTGT 1683

76 CCCAACAGAAAGGTTTGTATTAGAACTGTTTCTTAAATGATTGACTCTCTAA 17

1684 TCTTAAAAAATAAAAA 1699

16 AAAAAAATAAAAAA 1

RESULT 13

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 821)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DPG/Gazdar

cDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LCM1397 row: j column: 14

High quality sequence stop: 658.

FEATURES

Location/Qualifiers

1..821

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4637293"

/clone_lib="NIH_MGC_18"

/tissue_type="large cell carcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: lung; Vector: pOTB7; Site.1: XhoI; Site.2:

EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGGAG(G). Library constructed by Ling Hong in the

laboratory of Gerald M. Rubin (University of California,

Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and

Superscript II RT (Life Technologies). Note: this is a

NIH-MGC Library."

BASE COUNT 203 a 238 c 204 g 176 t

ORIGIN

Query Match 41.3%; Score 706; DB 12; Length 821;

Best Local Similarity 95.2%; Pred. No. 4.6e-119;

Matches 771; Conservative 0; Mismatches 35; Indels 4; Gaps 4;

26 GAGGTCAGACAGCCGCCAGCCGCTCTCGCAGCAGCAGTGAATAGTGTGTACCTCC 85

|||||


```

Qy 815 TACCAGCAGTGGCCACAATCCTGGCCACGCGGGCTTTCACCTGTGCTAATGAGAGTGTG 874
Db 421 TACCAGCAGTGGCCACAATCCTGGCCACGCGGGCTTTCACCTGTGCTAATGAGAGTGTG 480
Qy 875 CTGGAGACCTTAAGTATGAGCAGATGATGATGATGATGATGATGATGATGATGATGATG 934
Db 481 CTGGAGACCTTAAGTATGAGCAGATGATGATGATGATGATGATGATGATGATGATGATG 540
Qy 935 TTTGCTGTGGACGGGAGGCGCGCTGGGACAGACTCCTTTTCTGTATGTGATGGAGCAG 994
Db 541 TTTGCTGTGGACGGGAGGCGCGCTGGGACAGACTCCTTTTCTGTATGTGATGGAGCAG 600
Qy 995 GTATTCATGAAGTGGGTATGAGCAGTGTCTCTCCCTCCAGAGTTCCTGGCAGCAGCCG 1054
Db 601 GTATTCATGAAGTGGGTATGAGCAGTGTCTCTCCCTCCAGAGTTCCTGGCAGCAGCCG 660
Qy 1055 ATCAAGGACTATCACAGTTACATGCTACAGATTAGTAAGCACTCTGGAAGTAATGA 1113
Db 661 ATCAAGGACTATCACAGTTACATGCTACAGATTAGTAAG-CACTCTGGAAGTAATGA 718

RESULT 15
B0887284
LOCUS      B0887284      899 bp      mRNA      linear      EST 16-AUG-2002
DEFINITION AGENCOURT_8748467 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6331976
5', mRNA sequence.
ACCESSION  B0887284
VERSION     B0887284.1 GI:22279298
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus.
REFERENCE   1 (bases 1 to 899)
AUTHORS     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE       Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL     NIH-MGC http://mgi.nci.nih.gov/.
COMMENT     National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-r@mail.nih.gov
            Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
            Ph.D.
            cDNA Library Preparation: ResGen, Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LRAM13788 row: n column: 09
            High quality sequence stop: 637.
FEATURES    Location/Qualifiers
             1..899
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /clone="IMAGE:6331976"
             /clone_lib="NIH_MGC_130"
             /lab_host="DH10B (phage-resistant)"
             /note="Organ: otcysts; Vector: pCMV-SPORT6.1.cdb;
             Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
             Primer: Oligo dt. Average insert size 1.95 kb.
             Constructed by ResGen, Invitrogen Corp. Note: this is a
             NIH_MGC Library."
BASE COUNT  227 a 266 c 220 g 185 t 1 others
ORIGIN

```

Query Match 41.0%; Score 702; DB 14; Length 899;
 Best Local Similarity 90.4%; Pred. No. 2.4e-118;
 Matches 761; Conservative 0; Mismatches 80; Indels 1; Gaps 1;

```

Qy 357 ACTGGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCCTTCGATTGCTCC 416
Db 31 ACTGGGAGAGATACCAATACCGTCAGGACAGACCAAGAGTTCCTTCGATTGCTCC 90

```

```

Qy 417 CACGGGAGTTCCCTCTGGTGGAAAGTCCATGACCCACCCCTGCACCAACCCCTCAGCCAACA 476
Db 91 CACGGGAGTTCCCTCTGGTGGAAAGTCCATGACCCGCGCCCTGCACCAAGCTCAGCCAACA 150
Qy 477 AGCGGAGCCCGCCACTATGCTGGACATCCCTTCAGAGCCATGTAGTCTCACCATCCATA 536
Db 151 AGCGGAGCCCGCCACGATGCTGGACATCCCTTCAGAACCGTGTAGCTCACCATCCATA 210
Qy 537 CGATTTCAGTTGATTTCAGCACAAACCGAGCTCTTCGCAACCTTATTGGCCAGAGCTCAGGCC 596
Db 211 CCATTTCAGTTGATTTCAGCACAAACCGAGCTCTTCGCAACCTTATTGGCCAGAGCTCAGGCC 270
Qy 597 AGAATCAGCAGCAGACAGAGAGGTGTAAAACTGAAGAGAGTGAACCTTCTCCCTCGTGGC 656
Db 271 AGATTCAGCAGCAGACAGAGAGGTGTGAAGCTGAAGAGAGTGAACCTTCTCCCTCGTGGC 330
Qy 657 CTGGGTCACCTCTCTCCCTGATGACCTCTCTGCTTTAGATTGTAAGAAATCCCAATGTCAC 716
Db 331 CTGGGTCACCTCTCTCCCTGATGACCTCTGAGCTTCCAGCTTTAGAAATCCCAATGTCAC 390
Qy 717 CATTTCAGATCCCGCAGTGAACCCAGAGAGTGAACCTTCTGCGGAAAGGGGAACCTG 776
Db 391 CATTTCAGATCCCGCAGTGAACCCAGAGAGTGAACCTTCTGCGGAAAGGGGAACCTG 450
Qy 777 TGACTGAACCTCAGCTGGCACTCTCTGCGCAGCTCTCTTACCAGGAGTGGCCACAATCC 836
Db 451 TGACAGAGCTGAGCTGGCACTCTCTGCGCAGCTCTCTTATCAAGCAGTGCGCCACAATCC 510
Qy 837 TGGCCCAACGCGGGCTTTGACTGTCTAATCAGAGTGTCTTGGAGAGCCCTTAACCTGATGG 896
Db 511 TGGCCCAACGAGGCTTTGAGTGTCTAATGAAAGTGTCTTGGAGAGCCCTTAACCTGATGG 570
Qy 897 CACATGAGTATTGCCTTAAGTTTACCAAGTTGCTGGCTTTTGTGGACCGGAGGCC 956
Db 571 CACATGAGTATTGCCTTAAGTTTACCAAGTTGCTGGCTTTTGTGGACCGGAGGCC 630
Qy 957 GGCTGGGACAGACTCCTTTTCTGATGTGATGGAGCAGGTATTCATGAAGTGGGTATTG 1016
Db 631 TGCTGGGACAGACTCCTTTTCTGATGTGATGGAGCAGGTATTCATGAAGTGGGTATTG 690
Qy 1017 GCAGTGTGCTCTCCCTCCAGAAAGTTCTGGCAGCAGCCGATCAAGGACTATCAGATTACA 1076
Db 691 GCAGTGTGCTCTCCCTCCAGAAAGTTCTGGCAGCAGCCGATCAAGGACTATCAGATTACA 750
Qy 1077 TGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAA-GGATTGTCAATCCTGAGAAG 1135
Db 751 TGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAAAGGATTGTGAATCCCGAGAAG 810
Qy 1136 GCCACAGAGGAGCGCTAAACCTGTGAAGATCAAGGAGGAACCTGTGAGCGACATCACTTTT 1195
Db 811 GCCACAGGAGACACTAAACCTGTGAAGATCAAGGAGAGCGCTGTGAGCGACATCACTTTT 870
Qy 1196 CC 1197
Db 871 CC 872

```

Search completed: July 18, 2003, 09:46:46
 Job time : 2572 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: July 18, 2003, 09:48:28 ; Search time 0.872038 seconds
(without alignments)
881.929 Million cell updates/sec

Title: US-09-857-308-5
Perfect score: 37
Sequence: 1 LYQAVATI 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	89.2	426	2 S51016	sodium-translocati
2	32	86.5	865	2 T41685	probable gamma-ada
3	31	83.8	269	2 T31318	hypothetical prote
4	30	81.1	307	2 S55596	hypothetical prote
5	30	81.1	312	2 H71948	probable tetraacyl
6	30	81.1	410	2 H81918	probable sodium-tr
7	30	81.1	410	2 C81185	Na(+)-translocatin
8	30	81.1	413	2 AF0393	NADH2 dehydrogenas
9	29	78.4	91	2 B82893	hypothetical prote
10	29	78.4	143	2 F82292	hypothetical prote
11	29	78.4	233	2 S77776	triacylglycerol li
12	29	78.4	264	2 J04111	triacylglycerol li
13	29	78.4	299	2 F83795	ABC transporter (A
14	29	78.4	363	2 A81837	hypothetical prote
15	29	78.4	370	2 C71926	cag island protein
16	29	78.4	421	2 S26246	glutamate/aspartat
17	29	78.4	421	2 S26247	glutamate/aspartat
18	29	78.4	502	2 T16921	hypothetical prote
19	29	78.4	503	2 J05078	glutamate transpor
20	29	78.4	507	2 T34005	hypothetical prote
21	29	78.4	518	2 T29633	hypothetical prote
22	29	78.4	532	2 T23481	hypothetical prote
23	29	78.4	565	2 I38399	glutamate/aspartat
24	29	78.4	572	2 A55676	excitatory amino a
25	29	78.4	573	2 S28901	glutamate transpor
26	29	78.4	574	2 I37426	glutamate transpor
27	29	78.4	574	2 J04262	glutamate transpor
28	29	78.4	574	2 I38432	excitatory amino a
29	29	78.4	575	2 T29354	hypothetical prote

two-component sens
semaphorin B - mou
hypothetical prote
sensor protein evg
probable sensor fo
hypothetical prote
hypothetical prote
hypothetical prote
conserved hypotHet
hypothetical prote
hypothetical 22K p
probable shikimate
probable phenylace
germin-like protei
hemagglutinin - Ma

ALIGNMENTS

RESULT 1

S51016

sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) nqrB chain - Vibrio
N:Alternate names: NADH-ubiquinone oxidoreductase nqrB protein
C:Species: Vibrio alginolyticus
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 21-Jul-2000
C:Accession: S51016
R:Beattie, P.; Tan, K.; Bourne, R.M.; Leach, D.; Rich, P.R.; Ward, F.B.
FEBS Lett. 356, 333-338, 1994
A:Title: Cloning and sequencing of four structural genes for the Na(+)-translocating
A:Reference number: S51013; MUID:95104445; PMID:7805867
A:Accession: S51016
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-426 <BEA>
A:Cross-references: EMBL:Z37111; NID:g663268; PIDN:CAA85477.1; PID:g663271
C:Genetics:
A:Gene: nqrB
C:Keywords: NAD; oxidoreductase

Query Match 89.2%; Score 33; DB 2; Length 426;
Best Local Similarity 75.0%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
DB 38 LYEA VATV 45
||:||||:
|::|::|::|

RESULT 2

T41685

probable gamma-adaptin - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T41685
R:Wedler, H.; Duesterhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z21742
A:Accession: T41685
A:Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-865 <WED>
A:Cross-references: EMBL:AL117183; PIDN:CAB54865.1; GSPDB:GN00068; SPDB:SPCP1E11.06
A:Experimental source: strain 972h-; clone pl p1E11
C:Genetics:
A:Gene: SPDB:SPCP1E11.06
A:Map position: 3

Query Match 86.5%; Score 32; DB 2; Length 865;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

A: Experimental source: strain J99
C: Genetics:
A: Gene: lpxK
C: Superfamily: Escherichia coli probable tetracyclisaccharide 4'-kinase ynfH
C: Keywords: phosphotransferase

Query Match 81.1%; Score 30; DB 2; Length 312;
Best Local Similarity 62.5%; Pred. No. 40;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0

Qy 1 LYQAVATI 8
||:||||
Db 32 IYOCIAI 39

RESULT 6
H81918
probable sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) chain
C: Species: Neisseria meningitidis
C: Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C: Accession: H81918
R: Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.;
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; P
Nature 404, 502-506, 2000
A: Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis*
A: Reference number: A81775; MUID:20222556; PMID:10761919
A: Accession: H81918
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-410 <PAR>
A: Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84034.1; F
A: Experimental source: serogroup A, strain Z2491
C: Genetics:
A: Gene: nqrB: NMA0751
C: Keywords: NAD; oxidoreductase

Query Match 81.1%; Score 30; DB 2; Length 410;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0

Qy 1 LYQAVATI 8
||:||||
Db 26 LYEAAATI 33

RESULT 7
C81185
Na(+)-translocating NADH-quinone reductase, chain B NMB0568 [imported] - Neis
C: Species: Neisseria meningitidis
C: Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C: Accession: C81185
R: Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.;
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dough
Ric, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masiugnani, V.; Piz
Science 287, 1809-1815, 2000
A: Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappu
A: Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain
A: Reference number: A81000; MUID:20175755; PMID:10710307
A: Accession: C81185
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-410 <YET>
A: Cross-references: GB:AE002412; GB:AE002098; NID:g7225783; PIDN:AAF40996.1; F
A: Experimental source: serogroup B, strain MC58
C: Genetics:
A: Gene: NMB0568

Query Match 81.1%; Score 30; DB 2; Length 410;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0

Qy 1 LYQAVATI 8
||:||||

Db 26 LYEAAATI 33

RESULT 8

AF0393

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain B [imported] - Yersinia pestis (strain AF0393)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 03-Jun-2002
 C:Accession: AF0393
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; et al., M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell, Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AF0393
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-413 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC92474.1; PID:g15981175; GSPDB:GN00175
 C:Genetics:
 A:Gene: nqrB
 C:Keywords: oxidoreductase

Query Match 81.1%; Score 30; DB 2; Length 413;
 Best Local Similarity 75.0%; Pred. No. 54;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

Db 26 LYEAAATI 33

RESULT 9

B82893

hypothetical protein UU424.1 [imported] - Ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: B82893
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H. submitted to GenBank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mitochondrial genome
 A:Reference number: A82870
 A:Accession: B82893
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-91 <GLA>
 A:Cross-references: GB:AE002139; GB:AF222894; NID:96899405; PIDN:AAF30836.1; GSPDB:GN00175
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: UU424.1
 A:Genetic code: SGC3

Query Match 78.4%; Score 29; DB 2; Length 91;
 Best Local Similarity 87.5%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

Db 53 LLQAVATI 70

RESULT 10

F82292

hypothetical protein VC0697 [imported] - Vibrio cholerae (strain N15961 serogroup O1)
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: F82292
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; et al., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: F82292

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-143 <HEU>

A:Cross-references: GB:AE004155; GB:AE003852; NID:9655127; PIDN:AAF93862.1; GSPDB:GN00175
 A:Experimental source: serogroup O1; strain N15961; biotype El Tor

C:Genetics:

A:Gene: VC0697

A:Map position: 1

Query Match 78.4%; Score 29; DB 2; Length 143;

Best Local Similarity 75.0%; Pred. No. 30;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

Db 31 LYQAPATL 38

RESULT 11

S77776

triacylglycerol lipase (EC 3.1.1.3) - Mycoplasma capricolum (fragment)

N:Alternate names: lipase; protein MC073

C:Species: Mycoplasma capricolum

C:Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 21-Jul-2000

C:Accession: S77776

R:Bork, P.; Ouzounis, C.; Casari, G.; Schneider, R.; Sander, C.; Dolan, M.; Gilbert, M.O. Microbiol. 16, 955-967, 1995

A:Title: Exploring the Mycoplasma capricolum genome: a minimal cell reveals its physiology

A:Reference number: S77739; MUID:96059641; PMID:7476192

A:Accession: S77776

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-233 <BOR>

A:Cross-references: EMBL:Z33059; NID:9531610; PIDN:CAA83733.1; PID:94379137

A:Experimental source: AFCC 27343

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, July 1994

C:Comment: This enzyme, serine esterase, is widely distributed throughout animal

in fatty acids, and hydrolyses ester bonds of triacylglycerols to yield free fatty ac

C:Genetics:

A:Genetic code: SGC3

C:Superfamily: triacylglycerol lipase 1

C:Keywords: carboxylic ester hydrolase

Query Match 78.4%; Score 29; DB 2; Length 233;

Best Local Similarity 62.5%; Pred. No. 50;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

Db 165 IYQAIKTI 172

RESULT 12

JC4111

triacylglycerol lipase (EC 3.1.1.3) 3 - Mycoplasma mycoides subsp. mycoides

N:Alternate names: lipase

C:Species: Mycoplasma mycoides subsp. mycoides

C:Date: 23-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 07-Dec-1999

C:Accession: JC4111

R:Rawadi, G.; Lalanne, J.L.; Roulland-Dussoix, D.

Gene 158, 107-111, 1995

A:Title: Cloning and characterization of the lipase operon from Mycoplasma mycoides s

A:Reference number: JC4109; MUID:95309706; PMID:7789792

A:Accession: JC4111

A:Molecule type: DNA

A:Residues: 1-264 <RAW>

A:Cross-references: GB:U17036

C:Comment: This enzyme, serine esterase, is widely distributed throughout animal

in fatty acids, and hydrolyses ester bonds of triacylglycerols to yield free fatty ac

C:Genetics:

A:Genetic code: SGC3

C:Superfamily: triacylglycerol lipase 1
C:Keywords: carboxylic ester hydrolase
F:90-94/Region: conserved site

Query Match 78.4%; Score 29; DB 2; Length 264;
Best Local Similarity 62.5%; Pred. No. 58;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
:||||: ||
Db 197 IYQAIKTI 204

RESULT 13

ABC transporter (ATP-binding protein) nata [Imported] - Bacillus halodurans (strain C-12)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: F83795
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11059132
A:Accession: F83795
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-299 <STO>
A:Cross-references: GB:AP001511; GB:BA000004; NID:g10173727; PIDN:BA804885.1; GSPDB:GN00
C:Genetics:
A:Gene: nata
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 78.4%; Score 29; DB 2; Length 299;
Best Local Similarity 85.7%; Pred. No. 66;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVAT 7
I:|||||
Db 266 LFQAVAT 272

RESULT 14

AEI837
hypothetical protein all0245 [Imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AEI837
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AEI837
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-363 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA877769.1; PID:g17135223; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0245

Query Match 78.4%; Score 29; DB 2; Length 363;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 YQAVATI 8
||| |||
Db 120 YQATATI 126

RESULT 15

C71926

cag island protein - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: C71926

R:Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F
Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: C71926

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-370 <ARN>

A:Cross-references: GB:AE001482; GB:AE001439; NID:g4155018; PIDN:AAD06062.1; PID:g415

A:Experimental source: strain J99

C:Genetics:

A:Gene: cagH

Query Match 78.4%; Score 29; DB 2; Length 370;

Best Local Similarity 100.0%; Pred. No. 82;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YQAVAT 7

|||||

Db 241 YQAVAT 246

Search completed: July 18, 2003, 10:00:41

Job time : 2.87204 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2003, 07:39:42 ; Search time 4490 Seconds
(without alignments)
11090.178 Million cell updates/sec

Title: US-09-857-308-2

Perfect score: 1711

Sequence: 1 acgcgatcttgcctcaggc.....aaaaaaaaaaaaaaaaaaaa 1711

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:**

1: gb_ba:**

2: gb_htg:**

3: gb_in:**

4: gb_om:**

5: gb_ov:**

6: gb_pat:**

7: gb_ph:**

8: gb_pl:**

9: gb_pr:**

10: gb_ro:**

11: gb_sts:**

12: gb_sy:**

13: gb_un:**

14: gb_vi:**

15: em_ba:**

16: em_fun:**

17: em_hum:**

18: em_in:**

19: em_mu:**

20: em_om:**

21: em_or:**

22: em_ov:**

23: em_pat:**

24: em_ph:**

25: em_pl:**

26: em_ro:**

27: em_sts:**

28: em_un:**

29: em_vi:**

30: em_htg_hum:**

31: em_htg_inv:**

32: em_htg_other:**

33: em_htg_mus:**

34: em_htg_pln:**

35: em_htg_rod:**

36: em_htg_mam:**

37: em_htg_vrt:**

38: em_sy:**

39: em_htgo_hum:**

40: em_htgo_mus:**

41: em_htgo_other:**

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1691.8	98.9	2021	9	AF197954	AF197954 Homo sapi
2	1690.2	98.8	4261	9	AB018307	AB018307 Homo sapi
3	1688.6	98.7	2045	9	AK000329	AK000329 Homo sapi
4	1457.8	85.2	1481	9	AK000616	AK000616 Homo sapi
5	1340.8	78.4	2869	9	AF224759	AF224759 Homo sapi
6	1057	61.8	2616	10	BC031447	BC031447 Mus muscu
7	850.8	49.7	1046	6	AX013061	AX013061 Sequence
8	408	23.8	204153	9	AC074091	AC074091 Homo sapi
9	359	21.0	386	11	G25726	G25726 human STS E
10	356.8	20.9	173988	2	AC105160	AC105160 Mus muscu
11	356.8	20.9	232648	2	AC113276	AC113276 Mus muscu
12	352	20.6	155250	2	AC127781	AC127781 Rattus no
13	266.6	15.6	145971	2	AC096698	AC096698 Rattus no
14	253.2	14.8	145971	2	AC096698	AC096698 Rattus no
15	214.6	12.5	296	9	HS33F3F	HS33F3F H.sapiens C
16	208.4	12.2	338	9	HS181D8R	HS181D8R H.sapiens C
17	109	6.4	2954	9	AK001486	AK001486 Homo sapi
18	83.4	4.9	2949	9	AY028435	AY028435 Homo sapi
19	53	3.1	125020	9	AF429315	AF429315 Homo sapi
20	49.2	2.9	154609	2	AC118085	AC118085 Rattus no
21	48.8	2.9	155250	2	AC127781	AC127781 Rattus no
22	48.4	2.8	1340	6	E13530	E13530 3'untransla
23	48.4	2.8	1844	6	E16086	E16086 cDNA encodi
24	48.4	2.8	2502	6	AR103240	AR103240 Sequence
25	48.4	2.8	2502	6	BD000103	BD000103 Different
26	48.4	2.8	2502	9	D82347	D82347 Homo sapien
27	48.4	2.8	6804	9	AB009997	AB009997 Homo sapi
28	48.4	2.8	7251	9	AB018693	AB018693 Homo sapi
29	48.4	2.8	152306	9	AC013733	AC013733 Homo sapi
30	48.4	2.8	165020	9	AC073977	AC073977 Homo sapi
31	47.4	2.8	151113	9	AC093858	AC093858 Homo sapi
32	47.4	2.8	188000	2	AC007895	AC007895 Homo sapi
33	46.6	2.7	179968	9	AC026228	AC026228 Homo sapi
34	46.4	2.7	1015	3	AF020280	AF020280 Dictyoste
35	45.6	2.7	527	3	DDITND312	DDITND312 X53333 Slime mold
36	45.6	2.7	1369	10	MMTP1LR	MMTP1LR X53333 Mouse mRNA
37	45.6	2.7	160759	2	AC117082	AC117082 Dictyoste
38	45.4	2.7	33404	2	AC115684	AC115684 Dictyoste
39	45.2	2.6	1830	10	BC018241	BC018241 Mus muscu
40	45.2	2.6	185004	2	AC107844	AC107844 Mus muscu
41	45.2	2.6	195948	2	AC129326	AC129326 Mus muscu
42	44.8	2.6	1527	8	ATATN1	X92728 A.thaliana
43	44.8	2.6	293545	2	AC098276	AC098276 Rattus no
44	44.4	2.6	131665	2	CNS08CB9	AL845347 Oryza sat
45	44.4	2.6	173918	2	AC108353	AC108353 Rattus no

ALIGNMENTS

RESULT 1
AF197954

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AF197954 2021 bp mRNA linear PRI 05-JAN-2001
Homo sapiens adenocarcinoma antigen ART1 mRNA, complete cds.

AF197954.1 GI:11066257

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2021)

Nishizaka,S., Gomi,S., Harada,K., Oizumi,K., Itoh,K. and

Shichijo,S.

A new tumor-rejection antigen recognized by cytotoxic T lymphocytes

infiltrating into a lung adenocarcinoma
Cancer Res. 60 (17), 4830-4837 (2000)
20441578
MEDLINE
10987294
PUBMED
2 (bases 1 to 2021)
REFERENCE
Nishizaka,S., Gomi,S., Shichiho,S., Harada,K., Kawano,K. and
Itoh,K.
AUTHORS
Direct Submission
TITLE
Submitted (24-Oct-1999) Immunology, Kurume University School of
Medicine, Asahi-Machi 67, Kurume, Fukuoka 830-0011, Japan
JOURNAL
Location/Qualifiers
FEATURES
source
1..2021
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS
363..1607
/note="T cell recognized antigen"
/codon_start=1
/product="adenocarcinoma antigen ART1"
/protein_id="AAG28523.1"
/db_xref="GI:11066258"
translation="MNLYRYWGEIPISSQNRSGFDLLPREFLVEVDPLHOPSA
NKPKPTMLDIPSPESLT IHT IOLIOHNRRLRLN LATAQAOQOQT EGVKTESEPL
PSCGSPPLPDDLPLDCKNPNAPFQIRHSDPSDFYRGKGEPTVLSWHSQRLLIYQ
AVATLHAGFDCAVESVLETLTDVAHEYCLKFTKLRFVDRARLUGQTFPPDMEQ
VFHEVIGSVLSLQKFWQHRIDKDYHSMLOISKQSEYERIVRPERATEDAKPVKIK
EEPVSIDTFPVSLEADLASQSLPMLVLAQSERPPSNLEVEASPSQASAEVNAS
PLMNLAHVKMPQSEEGNVSCHGVLGSDVFEPMMSGMEAGIPQSPDDSDSYGSHS
TDSLNGSSPVFNQCKKRMKI"
BASE COUNT 510 a 523 c 489 g 499 t
ORIGIN
Query Match 98.9%; Score 1691.8; DB 9; Length 2021;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1693; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ACGGATCTTGCCCTCAGGCTCTCGAGGTCCAGACGCCGCCGCCGCTCTGCGACG 60
DB 23 ACGGATCTTGCCCTCAGGCTCTCGAGGTCCAGACGCCGCCGCCGCTCTGCGACG 82
QY 61 CAGCAGTGAATAGTGTGCTACTCTCTGCTCGGTTCAGGTCCAGACCTCCCGCTTCC 120
DB 83 CAGCAGTGAATAGTGTGCTACTCTCTGCTCGGTTCAGGTCCAGACCTCCCGCTTCC 142
QY 121 GCCTCCCTGACGCTCAGGCGACCTCAGACGCCCTGTGATTTGGCGCTCGCGCGGACG 180
DB 143 GCCTCCCTGACGCTCAGGCGACCTCAGACGCCCTGTGATTTGGCGCTCGCGCGGACG 202
QY 181 GTGACCGAGGAACCCCTGGAGGACCTTGGGATTCCTTGGGCTCCGCTGCTTCTTCG 240
DB 203 GTGACCGAGGAACCCCTGGAGGACCTTGGGATTCCTTGGGCTCCGCTGCTTCTTCG 262
QY 241 TGCTCTTTTGGGCAAGGATCTCACATTTATCAGTCTTTTGGACGACAGATGCTGGCA 300
DB 263 TGCTCTTTTGGGCAAGGATCTCACATTTATCAGTCTTTTGGACGACAGATGCTGGCA 322
QY 301 TTTGATAAATGTTTGTGAAGTGAAGACATATGGACAATGAATCTGCAAGATACTG 360
DB 323 TTTGATAAATGTTTGTGAAGTGAAGACATATGGACAATGAATCTGCAAGATACTG 382
QY 361 GGGAGAGATACCAATATATCAAGCCAGACCAAGAGTTCCTTCGATTTTGTCTCCACG 420
DB 383 GGGAGAGATACCAATATATCAAGCCAGACCAAGAGTTCCTTCGATTTTGTCTCCACG 442
QY 421 GGAGTTCGCTGTGTGGAAGTCCATGACCCACCCCTGCACCAACCCCTCAGCCACAAGCC 480
DB 443 GGAGTTCGCTGTGTGGAAGTCCATGACCCACCCCTGCACCAACCCCTCAGCCACAAGCC 502
QY 481 GAAGCCCCCCTACTATGCTGGACATCCCTTCAGAGCCATGTAGTCTCACCATTCATACGAT 540
DB 503 GAAGCCCCCCTACTATGCTGGACATCCCTTCAGAGCCATGTAGTCTCACCATTCATACGAT 562
QY 541 TCAGTTGATTCAGCACACCCAGCTCTTCGCAACCTTATTCGCACAGCTCAGGCCCGAGAA 600
|||||

563 TCAGTTGATTCAGCACAAACCGACGCTTTCGCAACCTTATTCGCACAGCTCAGGCCCGAGAA 622
QY 601 TCAGCAGCAGACAGAGAGGTGTAAGAACTGAAGAGAGTGAACCTCTTCCCTCGTCCCTGG 660
|||||
623 TCAGCAGCAGACAGAGAGGTGTAAGAACTGAAGAGAGTGAACCTCTTCCCTCGTCCCTGG 682
QY 661 GTCACCTCTCTCCCTGATGACCTCTCTGCTTTAGATTGTAAAGATCCCAATGACACCAT 720
|||||
683 GTCACCTCTCTCCCTGATGACCTCTCTGCTTTAGATTGTAAAGATCCCAATGACACCAT 742
QY 721 CCAGATCCGACAGTGAACCCAGAGAGTGAACCTTTATCTGCGGAAAGGGGAACTGTGAC 780
|||||
743 CCAGATCCGACAGTGAACCCAGAGAGTGAACCTTTATCTGCGGAAAGGGGAACTGTGAC 802
QY 781 TGAACCTGAGTGGCAGCTCTCTGCGCAGCTCTCTACAGCAGTGGGCAACATCTGCGC 840
|||||
803 TGAACCTGAGTGGCAGCTCTCTGCGCAGCTCTCTACAGCAGTGGGCAACATCTGCGC 862
QY 841 CCAGCGGGCTTTGACTGTGCTAATGAGAGTGTCTCTGGAGACCCCTAACTGATGTGGCACA 900
|||||
863 CCAGCGGGCTTTGACTGTGCTAATGAGAGTGTCTCTGGAGACCCCTAACTGATGTGGCACA 922
QY 901 TGAGTATTGCTTAAAGTTTACCAAGTTCCTGCTGCTTTGCTGTGACCCGGAGGCCGCT 960
|||||
923 TGAGTATTGCTTAAAGTTTACCAAGTTCCTGCTGCTTTGCTGTGACCCGGAGGCCGCT 982
QY 961 GGGACAGACTCTTTTCTGATGTGATGAGCAGGTATTCATGAAGTGGGTATTGGCAG 1020
|||||
983 GGGACAGACTCTTTTCTGATGTGATGAGCAGGTATTCATGAAGTGGGTATTGGCAG 1042
QY 1021 TGTGCTCTCTCCAGAAAGTTCTGCGCAGCACCGCATCAAGGACTATCACAGTTACATGCT 1080
|||||
1043 TGTGCTCTCTCCAGAAAGTTCTGCGCAGCACCGCATCAAGGACTATCACAGTTACATGCT 1102
QY 1081 ACAGATTAGTAAAGCACTCTCTGAAGATATGAAGGATTTGTCATCTGAGAGGCCAC 1140
|||||
1103 ACAGATTAGTAAAGCACTCTCTGAAGATATGAAGGATTTGTCATCTGAGAGGCCAC 1162
QY 1141 AGAGACGCTAAACCTGTAAGATCAAGGAGAACCTGTGAGCAGCATCACTTTTCCGT 1200
|||||
1163 AGAGACGCTAAACCTGTAAGATCAAGGAGAACCTGTGAGCAGCATCACTTTTCCGT 1222
QY 1201 CAGTGAGAGCTGGAGGCTGACCTTGTCTTGGAGACAGCTCACTGCTTATGGAGTCT 1260
|||||
1223 CAGTGAGAGCTGGAGGCTGACCTTGTCTTGGAGACAGCTCACTGCTTATGGAGTCT 1282
QY 1261 TGGGCTCAGACGCAACGCTTCCCATCTAACTGGAGTTGAAGCTTCAACACAGGCTTC 1320
|||||
1283 TGGGCTCAGACGCAACGCTTCCCATCTAACTGGAGTTGAAGCTTCAACACAGGCTTC 1342
QY 1321 AAGTGCAGAGTAAATGCTTCTCTTGGAACTGCGCCCATGTGAAATGGAGCTCA 1380
|||||
1343 AAGTGCAGAGTAAATGCTTCTCTTGGAACTGCGCCCATGTGAAATGGAGCTCA 1402
QY 1381 AGAAGTGAAGAGCAATGCTCTGGGCAATGCTGCTGGGAGTGAATGCTTTCGAGGA 1440
|||||
1403 AGAAGTGAAGAGCAATGCTCTGGGCAATGCTGCTGGGAGTGAATGCTTTCGAGGA 1462
QY 1441 GCCTATGTCAGGCATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAG 1500
|||||
1463 GCCTATGTCAGGCATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAG 1522
QY 1501 CTATGGTTCCCACTCCACTGACAGCCTCATGGGTCCTCCCTGTTTTCACACCGCTG 1560
|||||
1523 CTATGGTTCCCACTCCACTGACAGCCTCATGGGTCCTCCCTGTTTTCACACCGCTG 1582
QY 1561 CAAGAAGAGTGAAGAAATATATAAGAAAGAGGAGATGTTTGTTCAGACCTACT 1620
|||||
1583 CAAGAAGAGTGAAGAAATATATAAGAAAGAGGAGATGTTTGTTCAGACCTACT 1642
QY 1621 AGACCCCAACAGAAAGGTTTTTGTATTAGATCTGTTTCCCTTAAAAATTTGATTTGAC 1680
|||||
1643 AGACCCCAACAGAAAGGTTTTTGTATTAGATCTGTTTCCCTTAAAAATTTGATTTGAC 1702


```

Qy 1681 TGTTCTTAAAAAAA 1695
Db 1703 TGTTCCTAAACACAA 1717

RESULT 2
AB018307
LOCUS
DEFINITION Homo sapiens mRNA for KIAA0764 protein, complete cds.
ACCESSION AB018307
VERSION AB018307.1 GI:3882248
KEYWORDS
SOURCE
ORGANISM
Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
SK plus clone:hk04750.
Homo sapiens
Eukaryote; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Miyajima, N.,
Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.
Prediction of the coding sequences of unidentified human genes. XI.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
JOURNAL DNA Res. 5 (5), 277-286 (1998)
MEDLINE 99087487
REFERENCE 2 (bases 1 to 4261)
AUTHORS Ohara, O., Suyama, M., Nagase, T., Ishikawa, K. and Kikuno, R.
DIRECT SUBMISSION
TITLE Submitted (08-OCT-1998) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp. Tel: +81-438-52-3913,
Fax: +81-438-52-3914)
FEATURES
Location/Qualifiers
1..4261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hk04750"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
/dev_stage="adult"
1..4261
/gene="KIAA0764"
344..1588
/gene="KIAA0764"
/codon_start=1
/product="KIAA0764 protein"
/protein_id="BAA34484.1"
/db_xref="GI:3882248"
/translation="MNLQRYWGEIPISSQTNRSFDLLPREFLRVVHDPPLHOPSA
NKPPPTMLDIPSECSLTHTIQLIQRRLRLNLATAQONQOQTEGVKTESEPL
PSCPSPPLDLLPLDCKNPNAPFIRHSDPESDFYRGKGPVTELSWHSCRLLIYQ
AVATILAHGFDCANESVLETLTVAHEYCLKFTKLRLFAVDREARLGTPTPPVMEQ
VFEHVGISVLSCFKWHRIDYHSYMLQISKLSEYERIVNPEKATEDAKPVKIK
EFVSDISVSEELRADLSGDSLPMLGVAQSRRFSPNLEVEASPOASSAEVNAS
PLWNLAHVMEQESSEGVNMGVLSGVDFEPEPMGSMSEAGIPQSPDDSDSYGSHS
TDSLMSGSPVFVNRCKMRKI"
BASE COUNT 1171 a 966 c 897 g 1227 t
ORIGIN
Query Match 98.8%; Score 1690.2; DB 9; Length 4261;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1692; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 ACCGGATCTTGCTCAGGCTCTCGAGGTCCAGACGCGCCAGCCGCTCTGGGACG 60
Db 4 ACCGGATCTTGCTCAGGCTCTCGAGGTCCAGACGCGCCAGCCGCTCTGGGACG 63
Qy 61 CAGCAGTGAATAGTGGTACCTCTGCTCGGTTCCAGCTCCAGACCTCCCGCTTCC 120
Db 64 CAGCAGTGAATAGTGGTACCTCTGCTCGGTTCCAGCTCCAGACCTCCCGCTTCC 123

```

```

121 GGTGCCCTGAACGTGACGCGACCTCAGGACCCCTGTGATTTGGCGCCTGCGCCGCGGACC 180
124 GGCTGCCCTGAACGTGACGCGACCTCAGGACCCCTGTGATTTGGCGCCTGCGCCGCGGACC 183
181 GTGACCGAGAAACCCCTGAGGAGCTTGGGCATTCTCTTGGGCTCCGTGCTTCTTCG 240
184 GTGACCGAGAAACCCCTGAGGAGCTTGGGCATTCTCTTGGGCTCCGTGCTTCTTCG 243
241 TGTCTCTTTCGGGCAAGGATCTCAGTATATCAGTCTTTGACCCGACACAGAAATGCTTGCA 300
244 TGTCTCTTTCGGGCAAGGATCTCAGTATATCAGTCTTTGACCCGACACAGAAATGCTTGCA 303
301 TTTGATAAATGTTGTTGTAACCTTGAAGAGACATATGAGACAAATGAATCTGCAAGATACTG 360
304 TTTGATAAATGTTGTTGTAACCTTGAAGAGACATATGAGACAAATGAATCTGCAAGATACTG 363
361 GGGAGAGATACCAATATCATCAAGCCAGACCAACAGAAAGTTCCCTTCGATTTCCTCCACG 420
364 GGGAGAGATACCAATATCATCAAGCCAGACCAACAGAAAGTTCCCTTCGATTTCCTCCACG 423
421 GGAAGTCCGCTGCTGGTGAAGTCCATGACCCACCCCTGCACCAACCCCTCAGCCCAACAGCC 480
424 GGAAGTCCGCTGCTGGTGAAGTCCATGACCCACCCCTGCACCAACCCCTCAGCCCAACAGCC 483
481 GAAGCCCCCAGTATGCTGGACATCCCTCAGAGCCATGATGCTCACCATCCATAGAT 540
484 GAAGCCCCCAGTATGCTGGACATCCCTCAGAGCCATGATGCTCACCATCCATAGAT 543
541 TCAGTTGATTCACACACGCGCTCTTCGCAACCTTATTTGCCACAGCTCAGGCCAGAA 600
544 TCAGTTGATTCACACACGCGCTCTTCGCAACCTTATTTGCCACAGCTCAGGCCAGAA 603
601 TCAGCAGCAGACAGAGAGTGTAAAACTGAAGAGAGTGAACCTCTTCCCTCTGCTCCCTGG 660
604 TCAGCAGCAGACAGAGAGTGTAAAACTGAAGAGAGTGAACCTCTTCCCTCTGCTCCCTGG 663
661 GTCACCTCTCTCCCTGATGACCTCTGCTGCTTTAGATTGPAAGATCCCAATCCACAT 720
664 GTCACCTCTCTCCCTGATGACCTCTGCTGCTTTAGATTGPAAGATCCCAATCCACAT 723
721 CCAGATCCGCGCAGCAGTGACCCAGAGAGTGAATTTATCTGTTGGAAAGGGAACCTGTGAC 780
724 CCAGATCCGCGCAGCAGTGACCCAGAGAGTGAATTTATCTGTTGGAAAGGGAACCTGTGAC 783
781 TGAACCTCAGCTGGCAGCTCTGCTGGCAGCTCTTACCCAGCAGTGGCCACAACTCCTGGC 840
784 TGAACCTCAGCTGGCAGCTCTGCTGGCAGCTCTTACCCAGCAGTGGCCACAACTCCTGGC 843
841 CCACGCGGCTTTGACTGTGCTAATGAGAGTGTCTTGGAGACCCCTAACATGATGTGACACA 900
844 CCATGCGGCTTTGACTGTGCTAATGAGAGTGTCTTGGAGACCCCTAACATGATGTGACACA 903
901 TGAGTATTGCCTTAAGTTTACCAAGTTGCTGCTGCTTTGCTGTGGACCGGGAGCGCGCT 960
904 TGAGTATTGCCTTAAGTTTACCAAGTTGCTGCTGCTTTGCTGTGGACCGGGAGCGCGCT 963
961 GGCACAGACTCCCTTTTCTGATGTGATGGAGCAGGTATTCATGAAATGGTATTTGGCAG 1020
964 GGCACAGACTCCCTTTTCTGATGTGATGGAGCAGGTATTCATGAAATGGTATTTGGCAG 1023
1021 TGTGCTCTCCCTCCAGAAAGTTCTGGCAGCACCAGCATCAAGGACTATCACAGTTACATGCT 1080
1024 TGTGCTCTCCCTCCAGAAAGTTCTGGCAGCACCAGCATCAAGGACTATCACAGTTACATGCT 1083
1081 ACAGATTAGTAAAGCACTCTCTGAAGATATGAAGAGTTGTCAATCTCTGAGAGGCGAC 1140
1084 ACAGATTAGTAAAGCACTCTCTGAAGATATGAAGAGTTGTCAATCTCTGAGAGGCGAC 1143
1141 AGAGGAGCGCTAAACCTGTGAAGATCAAGGAGGAACCTGTGAGCGACATCACATTTTCTCT 1200
1144 AGAGGAGCGCTAAACCTGTGAAGATCAAGGAGGAACCTGTGAGCGACATCACATTTTCTCT 1203
1201 CAGTGAGAGCTGGAGGCTGACCTTGTCTTGGAGACCACTGCTCAGTCCCTATGGAGTGCT 1260

```

```
Db 1204 CAGTCAGAGCTGGAGGCTGACCTTGTCTGGAGACAGTCACTGCTATGGAGTCT 1263
Qy 1261 TGGGCTCAGAGCAAGCTTCCATCTAACCCTGAGGTGAAGCTTCCACACAGGCTTC 1320
Db 1264 TGGGCTCAGAGCAAGCTTCCATCTAACCCTGAGGTGAAGCTTCCACACAGGCTTC 1323
Qy 1321 AAGTGCAGAGTAATCTTCTCTCTTTGGAATCTGGCCCATGTGAAATGGAGCCTCA 1380
Db 1324 AAGTGCAGAGTAATCTTCTCTCTTTGGAATCTGGCCCATGTGAAATGGAGCCTCA 1383
Qy 1381 AGAAGTGAAGAAGCAATGCTCTGGCATGGTGTCTGGGAGTGAATGCTTTCGAGGA 1440
Db 1384 AGAAGTGAAGAAGCAATGCTCTGGCATGGTGTCTGGGAGTGAATGCTTTCGAGGA 1443
Qy 1441 GCCTATGTCAGGCATGAGTGAAGCTGCTTCCCTCAGAGCCCTGATGACTCAGATAGCAG 1500
Db 1444 GCCTATGTCAGGCATGAGTGAAGCTGCTTCCCTCAGAGCCCTGATGACTCAGATAGCAG 1503
Qy 1501 CTATGGTTCCTCCACTCCACTGACAGCCTCATGGGCTCTCCCTCTTTTCAACACGCGTG 1560
Db 1504 CTATGGTTCCTCCACTCCACTGACAGCCTCATGGGCTCTCCCTCTTTTCAACACGCGTG 1563
Qy 1561 CAAGAAGAGTGAAGAAATATATAAAGGAAAGAGGAGATGTTTGTCTCCAGACCTACT 1620
Db 1564 CAAGAAGAGTGAAGAAATATATAAAGGAAAGAGGAGATGTTTGTCTCCAGACCTACT 1623
Qy 1621 AGACCCCAACAGAAAGGTTTGTGATTAGAAATCTGTTTCTTAAATTTGATTGACTCC 1680
Db 1624 AGACCCCAACAGAAAGGTTTGTGATTAGAAATCTGTTTCTTAAATTTGATTGACTCC 1683
Qy 1681 TGTCTTTAAACAAAAA 1695
Db 1684 TGTCTTTAAACACAA 1698

RESULT 3
AK000329 2045 bp mRNA linear PRI 22-FEB-2000
LOCUS Homo sapiens cDNA FLJ20322 fis, clone HEP09455, highly similar to
DEFINITION AB018307 Homo sapiens mRNA for KIAA0764 protein.
ACCESSION AK000329
VERSION AK000329.1 GI:7020342
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP
clone:HEP09455.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,
Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T. and Sugano,S.
NEO human cDNA sequencing project
Unpublished
2 (bases 1 to 2045)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency). Location/Qualifiers
source I. .2045
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP09455"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone_lib="HEP"
/note="cloning vector pME18SFL3"
1. .2045
misc_feature
/note="highly similar to AB018307 Homo sapiens mRNA for
KIAA0764 protein"
BASE COUNT 529 a 526 c 489 g 501 t
ORIGIN
Query Match 98.7%; Score 1688.6; DB 9; Length 2045;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1691; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 ACGGATCTTCCTCAGGCTCTCGAGGTCCAGACAGCCGCCAGCCGCTCTGCGGACG 60
Db 32 ACGGATCTTCCTCAGGCTCTCGAGGTCCAGACAGCCGCCAGCCGCTCTGCGGACG 91
Qy 61 CAGCAGTGAATAGTGTGCTACCTCTGCTTCAGGTTCAGGTCACAGACTCCCGCTCTTCC 120
Db 92 CAGCAGTGAATAGTGTGCTACCTCTGCTTCAGGTTCAGGTCACAGACTCCCGCTCTTCC 151
Qy 121 GGCTGCCCTGAACGCTCAGGCGACCTCAGGACCCCTGTGTTGGCGCCTGCGCGCGGACG 180
Db 152 GGCTGCCCTGAACGCTCAGGCGACCTCAGGACCCCTGTGTTGGCGCCTGCGCGCGGACG 211
Qy 181 GTGACCGAGGAAACCCCTGGAGGACTTGGGCAATCTTGGGCTCGGCTGTCTTCTTCG 240
Db 212 GTGACCGAGGAAACCCCTGGAGGACTTGGGCAATCTTGGGCTCGGCTGTCTTCTTCG 271
Qy 241 TGCTCCTTCGGCAAGGATCTCATTATCAGTCTTTGACCGACACAGATGCTGGCA 300
Db 272 TGCTCCTTCGGCAAGGATCTCATTATCAGTCTTTGACCGACACAGATGCTGGCA 331
Qy 301 TTTGATAAATGTTTGTGTTGAAGACATATGGAACATGGAATCTGCAAGATACG 360
Db 332 TTTGATAAATGTTTGTGTTGAAGACATATGGAACATGGAATCTGCAAGATACG 391
Qy 361 GGGAGATACCAATATCATCAAGCCAGACCAACAGAGTTCCTTCGATTTGCTCCACG 420
Db 392 GGGAGATACCAATATCATCAAGCCAGACCAACAGAGTTCCTTCGATTTGCTCCACG 451
Qy 421 GGAGTTCCTGCTGGTGAAGTCCATGACCCCTGCACCAACCCCTCAGCACAACAGCC 480
Db 452 GGAGTTCCTGCTGGTGAAGTCCATGACCCCTGCACCAACCCCTCAGCACAACAGCC 511
Qy 481 GAAGCCCCCCTACTATGCTGGACATFCCCTCTCAGAGCCATGTAGTCTACCAATCCATGAT 540
Db 512 GAAGCCCCCCTACTATGCTGGACATFCCCTCTCAGAGCCATGTAGTCTACCAATCCATGAT 571
Qy 541 TCAGTTGATTTCAGCAACACGAGCTCTTGGCAACCTTATTGCCACAGCTCAGGCCCGAGAA 600
Db 572 TCAGTTGATTTCAGCAACACGAGCTCTTGGCAACCTTATTGCCACAGCTCAGGCCCGAGAA 631
Qy 601 TCAGCAGCAGACAGAGGTGTAAACCTGAAGAGAGTGAACCTCTTCCCTGCTGCCCTGG 660
Db 632 TCAGCAGCAGACAGAGGTGTAAACCTGAAGAGAGTGAACCTCTTCCCTGCTGCCCTGG 691
Qy 661 GTCACTTCCTCTCCTGATGACCTCTTGGCTTTAGATTGTAAAGATCCCAATGCCCAT 720
Db 692 GTCACTTCCTCTCCTGATGACCTCTTGGCTTTAGATTGTAAAGATCCCAATGCCCAT 751
Qy 721 CCAGATCCGCGACAGTCAAGCCAGAGAGTACTTTTATCGTGGGAAAGGGAACCTGTGAC 780
Db 752 CCAGATCCGCGACAGTCAAGCCAGAGAGTACTTTTATCGTGGGAAAGGGAACCTGTGAC 811
Qy 781 TGAAGTCAAGTGGCACTCTGCTGGGAGCTCTTACACAGCAGTGGCCACAATCTCTGGC 840
Db 812 TGAAGTCAAGTGGCACTCTGCTGGGAGCTCTTACACAGCAGTGGCCACAATCTCTGGC 871
```


QY 671 CTCCTGATGACCTCCTGCTTTAGATTGTAAGAAATCCCAATGACCAATTCAGATCCGG 730
Db 488 CTCCTGATGACCTCCTGCTTTAGATTGTAAGAAATCCCAATGACCAATTCAGATCCGG 547
QY 731 CACAGTGACCCAGAGAGTACTTTTATCGTGGGAAGGGGACCTGCTGACTCAACTCAGC 790
Db 548 CACAGTGACCCAGAGAGTACTTTTATCGTGGGAAGGGGACCTGCTGACTCAACTCAGC 607
QY 791 TGGCACTCCTGCTGCGCAGCTCCTTACCAGGAGTGGCCACAAATCCTGCGCCACCGGCG 850
Db 608 TGGCACTCCTGCTGCGCAGCTCCTTACCAGGAGTGGCCACAAATCCTGCGCCACCGGCG 667
QY 851 TTGTGACTGCTTAATGAGAGTCTCTGGAGACCCCTAACTGATGTCGACATGATGCTTGC 910
Db 668 TTGTGACTGCTTAATGAGAGTCTCTGGAGACCCCTAACTGATGTCGACATGATGCTTGC 727
QY 911 CTTAAGTTTACCAAGTCTGCTGGTTTCTGCTGGACCGGAGGCGGCTGGGACAGACT 970
Db 728 CTTAAGTTTACCAAGTCTGCTGGTTTCTGCTGGACCGGAGGCGGCTGGGACAGACT 787
QY 971 CTTTTTCTGATGATGGAGCAGGTATTCCATGAAGTGGGTATTGGCAGTGTGCTCTCC 1030
Db 788 CTTTTTCTGATGATGGAGCAGGTATTCCATGAAGTGGGTATTGGCAGTGTGCTCTCC 847
QY 1031 CTCGAGAGTCTGCGCAGACCGCATCAAGGACTATCACAGTTACATGCTACAGATTAGT 1090
Db 848 CTCGAGAGTCTGCGCAGACCGCATCAAGGACTATCACAGTTACATGCTACAGATTAGT 907
QY 1091 AAGCAACTCTCTGAAGATATGAAGGATTTGCTCAATCTGAGAAGGCCACAGAGACGCT 1150
Db 908 AAGCAACTCTCTGAAGATATGAAGGATTTGCTCAATCTGAGAAGGCCACAGAGACGCT 967
QY 1151 AAACCTGTGAAGATCAAGGAGGAACTGTGAGCAGCATCACTTTTCTGTGCTGAGGAG 1210
Db 968 AAACCTGTGAAGATCAAGGAGGAACTGTGAGCAGCATCACTTTTCTGTGCTGAGGAG 1027
QY 1211 CTGGAGGCTGACCTGCTCTGAGACACGACCTGCTGAGGAGTGTGCTGAGGAGTGTG 1270
Db 1028 CTGGAGGCTGACCTGCTCTGAGACACGACCTGCTGAGGAGTGTGCTGAGGAGTGTG 1087
QY 1271 AGCGAAGCTCTCCATCTAACTGAGGTTGAAGCTTCCACACAGGCTTCAAGTGCAGAG 1330
Db 1088 AGCGAAGCTCTCCATCTAACTGAGGTTGAAGCTTCCACACAGGCTTCAAGTGCAGAG 1147
QY 1331 GFAAATGCTCTCTCTTTGGAATCTGCGCCATGTGAAATGAGGCTTCAAGAAAGTGA 1390
Db 1148 GFAAATGCTCTCTCTTTGGAATCTGCGCCATGTGAAATGAGGCTTCAAGAAAGTGA 1207
QY 1391 GAAGCAATGCTCTGCGCATGCTGCTGGGAGTGTGCTGGGAGTGTGCTGGGAGGCTATG 1450
Db 1208 GAAGCAATGCTCTGCGCATGCTGCTGGGAGTGTGCTGGGAGTGTGCTGGGAGGCTATG 1267
QY 1451 GGCATGAGTGAAGCTGCGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGTATGTTCC 1510
Db 1268 GGCATGAGTGAAGCTGCGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGTATGTTCC 1327
QY 1511 CACTCCACTGACAGCCTCATGGGCTCTCCCTGTTTCAACAGGCTTCAAGAAAGAGG 1570
Db 1328 CACTCCACTGACAGCCTCATGGGCTCTCCCTGTTTCAACAGGCTTCAAGAAAGAGG 1387
QY 1571 ATGAGGAAATATATAAGGAAAGAGGAGATGTTTTCAGAGCCTACTAGACCCACA 1630
Db 1388 ATGAGGAAATATATAAGGAAAGAGGAGATGTTTTCAGAGCCTACTAGACCCACA 1447
QY 1631 GAAAAGGTTTTGTTATAGAA 1651
Db 1448 GAAAAGGTTTTGTTATATAA 1468

RESULT 5
AF224759
LOCUS
DEFINITION

2869 bp mRNA linear PRI 07-JAN-2001
Homo sapiens adenocarcinoma antigen ART1/P17 mRNA, complete cds.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AF224759
AF224759.1 GI:12043739

Homo sapiens
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 2869)
Nishizaka, S., Gomi, S., Harada, K., Oizumi, K., Itoh, K. and
Shichijo, S.

TITLE
A new tumor-rejection antigen recognized by cytotoxic T lymphocytes
infiltrating into a lung adenocarcinoma

JOURNAL
MEDLINE

PUBMED
10987294

REFERENCE

2 (bases 1 to 2869)

Nishizaka, S., Gomi, S., Harada, K., Oizumi, K., Itoh, K. and
Shichijo, S.

Direct Submission

Submitted (14-JAN-2000)

Immunology, Kurume University School of
Medicine, Asahi-machi 67, Kurume, Fukuoka 830-0011, Japan

Location/Qualifiers

FEATURES
source

1..2869

/organism="Homo sapiens"

/db_xref="taxon:9606"

/tissue_type="adenocarcinoma"

1198..2436

/codon_start=1

/product="antigen ART1/P17"

/protein_id="AAG47636.1"

/db_xref="GI:12043740"

/translation="MLRYWGEIPISSTQNRSSFDLLPREFRLVEVHPDPLHPSANK
PKPMTLDIPSEPCSLTHTLIQHNRLNLATAQOQQTGKVESEPLPS
KPGSPPLDPLDLCKNPAPFOIRHSDPESDFVKGPEVTELSWISCRQLLYQAV
ATILAHAGFCANESVLETLDVAHEVCLKTKLLRFADREARLGPDPFDMQVFE
HEVGISVLSQKFWHRIKDYHSLMQLSKQSEERIVNPEKATADAPVKKEE
PVSDITFPVSEELADLASQSLPMGVIGASERFSPNLEVASPQASSAEFVNASPL
WNLARVMPEQSEEGNVSGHVLGSDVFEFPMGSMSEAGIQSPDDSDSYSGSHST
SLMGSPVFNQRCRKRKI"

BASE COUNT 769 a 673 c 655 g 772 t

ORIGIN

Query Match 78.4%; Score 1340.8; DB 9; Length 2869;

Best Local Similarity 99.5%; Pred. No. 0;

Matches 1345; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 344 AATCTGCAAGATCTGGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCC 403

Db 1195 AAAATGTTGAGATCTGGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCC 1254

QY 404 TTGATTTGCTCCACGGAGTTCCTGCTGGTGGAGTCCATGACCCACCTGACCAA 463

Db 1255 TTGATTTGCTCCACGGAGTTCCTGCTGGTGGAGTCCATGACCCACCTGACCAA 1314

QY 464 CCCTCAGCCACACAGCCGAGCCGCCCTACTATGCTGGACATCCCTCAGAGCATGTAGT 523

Db 1315 CCCTCAGCCACACAGCCGAGCCGCCCTACTATGCTGGACATCCCTCAGAGCATGTAGT 1374

QY 524 CTCACCATCCATGATTCAGTTCAGTTCAGCAGACACCGACCTCTTCGCAACCTTATGCC 583

Db 1375 CTCACCATCCATGATTCAGTTCAGTTCAGCAGACACCGACCTCTTCGCAACCTTATGCC 1434

QY 584 ACAGCTCAGGCGCCAGAAATCAGCAGCAGACAGAGGTGTAAACCTGAAGAGAGTGAACCT 643

Db 1435 ACAGCTCAGGCGCCAGAAATCAGCAGCAGACAGAGGTGTAAACCTGAAGAGAGTGAACCT 1494

QY 644 CTTCCCTGCTGCTGGTGGTCCCTCTCCCTGATGACCTCCCTGCTTGTAGTTGTAAG 703

Db 1495 CTTCCCTGCTGCTGGTGGTCCCTCTCCCTGATGACCTCCCTGCTTGTAGTTGTAAG 1554

QY 704 AATCCCAATGACCATTCAGATCCGCGCACAGTGCACCGAGAGTGACCTTTTATCGTGGG 763

Db 1555 AATCCCAATGACCATTCAGATCCGCGCACAGTGCACCGAGAGTGACCTTTTATCGTGGG 1614

QY	764	AAAGGGAACTGTGACTGAACCTCAGCTGGCAGCTCTCTCGCAGCTCTCTTACCAGCA	823	QY	1604	TTTTGTCCAGACCTACTAGACCCCAACGAAAGGTTTTTGTATTAGAAATCTGTTCTCTTA	1663
Db	1615	AAAGGGAACTGTGACTGAACCTCAGCTGGCAGCTCTCTCGCAGCTCTCTTACCAGCA	1674	Db	2455	TTTTGTCCAGACCTACTAGACCCCAACGAAAGGTTTTTGTATTAGAAATCTGTTCTCTTA	2514
QY	824	GTGGCCACAACTCTCGGCCACGGCGCTTACGTGTCTAATGAGAGTGTCTCGAGACC	883	QY	1544	GTTTTCAACGAGCCTCAAGAGAGGATGAGGAAATATAAAGGAAAGAGGAGATG	1603
Db	1675	GTGGCCACAACTCTCGGCCACGGCGCTTACGTGTCTAATGAGAGTGTCTCGAGACC	1734	Db	2395	GTTTTCAACGAGCCTCAAGAGAGGATGAGGAAATATAAAGGAAAGAGGAGATG	2454
QY	884	CTAACTGATGGGCACATGAGTATTGCTTAAGTTTACCAAGTTGCTCGCTTTGCTGTG	943	QY	1604	TTTTGTCCAGACCTACTAGACCCCAACGAAAGGTTTTTGTATTAGAAATCTGTTCTCTTA	1663
Db	1735	CTAACTGATGGGCACATGAGTATTGCTTAAGTTTACCAAGTTGCTCGCTTTGCTGTG	1794	Db	2455	TTTTGTCCAGACCTACTAGACCCCAACGAAAGGTTTTTGTATTAGAAATCTGTTCTCTTA	2514
QY	944	GACCGGAGGCCCGCTGGGACAGACTCTTTTCTGATGTGATGGAGCAGGTATTCAT	1003	QY	1664	AAATGATTGACTCTGTTCTTAAAAAAA	1695
Db	1795	GACCGGAGGCCCGCTGGGACAGACTCTTTTCTGATGTGATGGAGCAGGTATTCAT	1854	Db	2515	AAATGATTGACTCTGTTCTTAAAAACA	2546
QY	1004	GAAGTGGGTATTGGCAGTGTCTCTCCCTCAGAACTCTCGGACAGCCGATCAAGGAC	1063				
Db	1855	GAAGTGGGTATTGGCAGTGTCTCTCCCTCAGAACTCTCGGACAGCCGATCAAGGAC	1914				
QY	1064	TATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAAATATGAAGGATTGTC	1123				
Db	1915	TATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAAATATGAAGGATTGTC	1183				
QY	1124	AATCCTGAGAGGCCACAGAGGACGCTAAACCTGTGAAGATCAAGGAGGAACTGTGAGC	1183				
Db	1975	AATCCTGAGAGGCCACAGAGGACGCTAAACCTGTGAAGATCAAGGAGGAACTGTGAGC	2034				
QY	1184	GACATCACTTTTCTGTCAGTGAGGAGCTGGAGCTGACCTTCTCTTGGAGACCACTCA	1243				
Db	2035	GACATCACTTTTCTGTCAGTGAGGAGCTGGAGCTGACCTTCTCTTGGAGACCACTCA	2094				
QY	1244	CTGCCTATGGAGTGTCTGGGCTCAGAGCGAAGCTTCCCATCTAACTCGAGGAGTTGAA	1303				
Db	2095	CTGCCTATGGAGTGTCTGGGCTCAGAGCGAAGCTTCCCATCTAACTCGAGGAGTTGAA	2154				
QY	1304	GCTTACACAGGCTTCAAGTGCAGAGGTAAATGCTTCTCTCTTTGGAATCTGGCCCAT	1363				
Db	2155	GCTTACACAGGCTTCAAGTGCAGAGGTAAATGCTTCTCTCTTTGGAATCTGGCCCAT	2214				
QY	1364	GTGAAATGAGCCTCAAGAAAGTGAGAGGCAATGTCTCTGGGCATGTGTCTGGGC	1423				
Db	2215	GTGAAATGAGCCTCAAGAAAGTGAGAGGCAATGTCTCTGGGCATGTGTCTGGGC	2274				
QY	1424	AGTGATGCTCTCGAGGAGCCTATGTCAGGCATGAGTGAAGTGGGATTCTCTCAGAGCCT	1483				
Db	2275	AGTGATGCTCTCGAGGAGCCTATGTCAGGCATGAGTGAAGTGGGATTCTCTCAGAGCCT	2334				
QY	1484	GATGACTCAGATGACGATATGGTTCACCTCCACTCCAGCCCTCATGGGTCTCCCT	1543				
Db	2335	GATGACTCAGATGACGATATGGTTCACCTCCACTCCAGCCCTCATGGGTCTCCCT	2394				
QY	1544	GTTTTCAACGAGCCTCAAGAGAGGATGAGGAAATATAAAGGAAAGAGGAGATG	1603				
Db	2395	GTTTTCAACGAGCCTCAAGAGAGGATGAGGAAATATAAAGGAAAGAGGAGATG	2454				
QY	1604	TTTTGTCCAGACCTACTAGACCCCAACGAAAGGTTTTTGTATTAGAAATCTGTTCTCTTA	1663				
Db	2455	TTTTGTCCAGACCTACTAGACCCCAACGAAAGGTTTTTGTATTAGAAATCTGTTCTCTTA	2514				
QY	1664	AAATGATTGACTCTGTTCTTAAAAAAA	1695				
Db	2515	AAATGATTGACTCTGTTCTTAAAAACA	2546				
RESULT 6							
BC031447							
LOCUS							
DEFINITION							
IMAGE:4502314, mRNA, complete cds.							
ACCESSION							
BC031447							
VERSION							
BC031447.1							
GI:21594424							
KEYWORDS							
SOURCE							
ORGANISM							
house mouse.							
Mus musculus							
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							
1 (bases 1 to 2616)							
Direct Submission							
Submitted (06-JUN-2002) National Institutes of Health, Mammalian							
Gene Collection (MGC), Cancer Genomics Office, National Cancer							
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,							
USA							
NIH-MGC Project URL: http://mgc.nci.nih.gov							
Contact: MGC help desk							
Email: cgapbs-re@mail.nih.gov							
Tissue Procurement: The Cepko Laboratory							
cDNA Library Preparation: Life Technologies, Inc.							
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)							
DNA Sequencing by: Baylor College of Medicine Human Genome							
Sequencing Center							
Center code: BCM-HGSC							
Web site: http://www.hgsc.bcm.tmc.edu/cdna/							
Contact: amg@bcm.tmc.edu							
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,							
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,							
Richards, S., Gibbs, R.A.							
Clone distribution: MGC clone distribution information can be found							
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov							
Series: IRAP Plate: 31 Row: k Column: 15							
This clone was selected for full length sequencing because it							
passed the following selection criteria: Hexamer frequency ORF							
analysis, Genomescan gene prediction, similarity but not identity							
to protein.							
FEATURES							
source							
1. .2616							
/organism="Mus musculus"							
/db_xref="taxon:10090"							
/clone="MGC:25489 IMAGE:4502314"							
/tissue_type="Eye, retina, mouse strain C57Bl/6"							
/clone_lib="NIH_MGC_94"							
/lab_host="DH10B"							
/note="vector: pCMV-SPORT6"							
267. .1505							
/codon_start=1							
/product="RIKEN cDNA 2610524B01 gene"							
/protein_id="AAH31447.1"							
/db_xref="GI:21594425"							
/db_xref="LocusID:72195"							
/translation="MLRYWGEIPIPSGQTNRRSFDLLPRFLRVVEHDPPLHPSANK							
PKPTMLDIPSEPCSLTIHTLIQHNRRLSLIATQTSQQTGKVAESEPPLPS							
CPGSPPLDPLDCKPNAPFOIRHSDPESDFYRGKGPVTELSWHSCROLLYQAV							
ATLHATGFCANESVLETLTDVAHEYCLKFTKLRFAYDREALGTPPDVMEQVF							
HEVGISVLSQNFQWHRINDYHYMLQISQJSEERYIVNPEKATEDPKPKREE							
VNSDITPPVEEELADIASGDSLPIGLVQAQSERFSPNLEVASPOPAEYNASPL							
WNLAHVMEPESEEGVNSAGHVLGSDVFEEPMGSMSEAGLPQSPDDSDSYGSHSTD							
SLMGSSPVFNQRCRKRKI"							
BASE COUNT							
ORIGIN							
651 a 682 c 620 g 663 t							
Query Match							
Best Local Similarity 61.8%; Score 1057; DB 10; Length 2616;							
Matches 1129; Conservative 0; Mismatches 120; Indels 0; Gaps 0;							
QY	344	AAATGCAAGAGATCTGGGAGAGATACCAATATCATCAAGCCAGACACAGAGTTC	403	QY	344	AAATGCAAGAGATCTGGGAGAGATACCAATATCATCAAGCCAGACACAGAGTTC	403
Db	264	AAATGCTGAGGTACTGGGAGAGATACCAATATCGTCAGAGACACAGAGTTC	323	Db	264	AAATGCTGAGGTACTGGGAGAGATACCAATATCGTCAGAGACACAGAGTTC	323
QY	404	TTTCATGCTGCCACCGGAGTTCGCTGTGGTGAAGTCCATGATCCACCCCTGCACAA	463	QY	404	TTTCATGCTGCCACCGGAGTTCGCTGTGGTGAAGTCCATGATCCACCCCTGCACAA	463
Db	324	TTTCATGCTGCCACCGGAGTTCGCTGTGGTGAAGTCCATGATCCACCCCTGCACAA	383	Db	324	TTTCATGCTGCCACCGGAGTTCGCTGTGGTGAAGTCCATGATCCACCCCTGCACAA	383
QY	464	CCCTCAGCCAAACAGCCGAAAGCCGCCCACTATGCTGGACATCCCTCAGAGCCATG	523	QY	464	CCCTCAGCCAAACAGCCGAAAGCCGCCCACTATGCTGGACATCCCTCAGAGCCATG	523

QY 600 ATCAGCAGCAGACAGAGGTTAAAAAAGTGAAGAGAGTGAACTTCTTCCCTCGGCCCTG 659
Db 646 ATCAGCAGCAGACAGAGGTTAAAAAAGTGAAGAGAGTGAACTTCTTCCCTCGGCCCTG 705
QY 660 GGTCACTTCTCTCCCTGATGACCTCTGCTTTAGATTTGAAGAATCCCAATGACCAT 719
Db 706 GGTCACTTCTCTCCCTGATGACCTCTGCTTTAGATTTGAAGAATCCCAATGACCAT 765
QY 720 TCCAGATCCGCGACAGTGACCCAGAGAGTGACTTTTATCGTGGGAAAGGGGAACTGTGA 779
Db 766 TCCAGATCCGCGACAGTGACCCAGAGAGTGACTTTTATCGTGGGAAAGGGGAACTGTGA 825
QY 780 CTGAACCTCAGCTGCGACCTCTGTCGGAGCTCTCTACCG-AGGCAGTGGCCACAACTCTG 838
Db 826 CTGAACCTCAGCTGCGACCTCTGTCGGAGCTCTCTACCGAGCGAGTGGCCACAACTCTG 885
QY 839 GCCCAGC-CGGGCTTTGACTGTGCTAATGAGAGTGCTCGG-AGACCTTAACCT- -GATG 893
Db 886 GCCAACCGCGGCTTTGACTGTGCTAATGAGAGTGCTCGGAGAGCCCTTAACCTGATGT 945
QY 894 TGGCACATGATGATTTG--CCTTAAGTTTACCAAGTGTGCTGCG---TTTTGCTGTGGACCG 948
Db 946 TGGCACATGATGATTTGCTTAAAGTTTACCAAGTTTGTGCGCTTTTGTGCTGTGAGCG 1005
QY 949 GGAGCGCGCGTGGGAGAGCTCTCTTT 976
Db 1006 GGAAGCGCGGTGGGAGAGACTTCTCTTT 1033

RESULT 8
AC074091
DEFINITION Homo sapiens BAC clone RP11-158113 DNA linear PRI 10-JAN-2002
ACCESSION AC074091
VERSION AC074091.6 GI:17975414
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 204153)
AUTHORS Nguyen, C., Haglund, K., Dixon, R. and Doeber, A.
TITLE The sequence of Homo sapiens BAC clone RP11-158113
JOURNAL Unpublished (2002)
REFERENCE 3 (bases 1 to 204153)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 204153)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 204153)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 204153)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT

On Dec 21, 2001 this sequence version replaced gi:16924152.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0158113

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RPC11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frangen, E., Taten, M., Catanese, J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-158113; actual end is at base position 204153 of RP11-158113.

A PCR only region exists between bases 64163 to 64357. A transposon exists in this clone that has been deleted from the finished sequence. This transposon would insert after base 38173.

FEATURES

source

Location/Qualifiers
1. .204153
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-158113"
/clone_lib="RPC1-11"
464
/notes="match to EST AI370811 (NID:g4149564) ta58c03.xl"
misc_feature
492..800
/note="match to EST AA649214 (NID:g2575643) ns31f01.s1"
misc_feature
492..800
/note="match to EST BF516222 (NID:g11601401)"
misc_feature
497..800
/note="match to EST BI597621 (NID:g15490560)"
misc_feature
498..962
/note="match to EST AI370811 (NID:g4149564) ta58c03.xl"
misc_feature
500..800
/note="match to EST AV734630 (NID:g10852175)"
misc_feature
500..800
/note="match to EST BE881448 (NID:g10330224)"
misc_feature
500..800
/note="match to EST BG215442 (NID:g13741463)"
misc_feature
500..752
/note="match to EST AA594514 (NID:g2409864) n194f03.s1"
repeat_region
500..520

Primer B: TGACAGCCTCATGGGGTC
STS size: 130
PCR Profile:

Presoak:

Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:

Protocol:

Template: 10 ng
Primer: each 5 pm
dNTPs: each 4 mM
Taq Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:

MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 9.3

Derived from dbEST (genbank accession T48604).

Location/Qualifiers

1. .386
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="155.8 cR from top of Chr2 linkage group"
14. .143
14. .143

primer_bind

primer_bind complement(126. .143)

BASE COUNT 93 a 104 c 84 g 104 t 1 others

ORIGIN

Query Match 21.0%; Score 359; DB 11; Length 386;

Best Local Similarity 99.5%; Pred. No. 2.2e-91;

Matches 370; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY	1291	CCTGGAGGTTGAAGCTTCA-CCACAGGCTTCAAGTGCAGAGTAAATGCTTCCTCTTT	1349
Db	372	CCTGGAGGTTGAAGCTTCAACCCACAGGCTTCAAGTGCAGAGTAAATGCTTCCTCTTT	313
QY	1350	GGAATCTGGCCCATGTGAAATGGAGCCTCAAGAAAGTGGAAGCAATGCTCTCTGGGC	1409
Db	312	GGAATCTGGCCCATGTGAAATGGAGCCTCAAGAAAGTGGAAGCAATGCTCTCTGGGC	253
QY	1410	ATGGTGTGCTGGGCAGTGATGCTTCGAGGAGCCTATGTCAGGCATGAGTGAAGCTGGGA	1469
Db	252	ATGGTGTGCTGGGCAGTGATGCTTCGAGGAGCCTATGTCAGGCATGAGTGAAGCTGGGA	193
QY	1470	TTCTCTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTCCCACTCCACTGCAGAGCTCA	1529
Db	192	TTCTCTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTCCCACTCCACTGCAGAGCTCA	133
QY	1530	TGGGGTCTCCCTGTTTTCACACAGCGCTGCAAGAGAGGATGAGGAAATATAAAGG	1589
Db	132	TGGGGTCTCCCTGTTTTCACACAGCGCTGCAAGAGAGGATGAGGAAATATAAAGG	73
QY	1590	AAAGAGGAGGATGTTTGTTCAGACCTACTAGACCCACAGAAAAGGTTTGTATTAG	1649
Db	72	AAAGAGGAGGATGTTTGTTCAGACCTACTAGACCCACAGAAAAGGTTTGTATTAG	13
QY	1650	AATCTGTTTCCT 1661	
Db	12	AATCTGTTTCCT 1	

RESULT 10

AC105160

LOCUS

DEFINITION

AC105160

ACCESSION

AC105160 173988 bp DNA linear HTG 21-AUG-2002
Mus musculus clone RP24-555E12, WORKING DRAFT SEQUENCE, 3 ordered
pieces.
AC105160

VERSION

HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS

house mouse.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 173988)

Birren, B., Nusbaum, C. and Lander, E.

Mus musculus, clone RP24-555E12

Unpublished

2 (bases 1 to 173988)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,

Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Grinde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,

Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,

MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,

McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,

Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,

Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,

Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,

Seaman, S., Sever, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (26-Dec-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 173988)

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,

Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,

Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,

Cooke, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S.,

Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,

Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,

Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,

Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,

Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,

McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,

Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,

O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,

Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,

Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Sever, P.,

Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,

Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,

Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Aug 21, 2002 this sequence version replaced g1:20128245.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L18761

Center clone name: 555_E_12

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

LOCUS	AC113276	232648 bp	DNA	linear	HTG 24-AUG-2000
DEFINITION	Mus musculus clone RP23-383E21, WORKING DRAFT SEQUENCE, 6 ordered pieces.				
ACCESSION	AC113276				
VERSION	AC113276.2	GI:22475032			
KEYWORDS	HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 232648)				
TITLE	Birren,B., Nusbaum,C. and Lander,E.				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 232648)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Laococque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talanas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (28-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
REFERENCE	3 (bases 1 to 232648)				
AUTHORS	Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gerold,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talanas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (28-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
REFERENCE	3 (bases 1 to 232648)				
AUTHORS	Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gerold,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talanas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT	On Aug 24, 2002 this sequence version replaced gi:18997558. All repeats were identified using RepeatMasker: Smit,A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html				
TITLE	Genome Center				
JOURNAL	Center: Whitehead Institute/ MIT Center for Genome Research				
COMMENT	Center code: WIBR Web site: http://www-seq.wi.mit.edu Contact: sequence_submissions@genome.wi.mit.edu ----- Project Information Center project name: L22709				

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 155250)

Worley,K.C.

Direct Submission

Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZSG
Center clone name: CH230-46303
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 88645 bases at least Q40
Consensus quality: 95723 bases at least Q30
Consensus quality: 100316 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1151: contig of 1151 bp in length
* 1152 1251: gap of unknown length
* 1252 2566: contig of 1315 bp in length
* 2567 2667: gap of unknown length
* 2667 4223: contig of 1557 bp in length
* 4224 4324: gap of unknown length
* 4324 5405: contig of 1082 bp in length
* 5406 5505: gap of unknown length
* 5506 6569: contig of 1064 bp in length
* 6570 6669: gap of unknown length
* 6670 8299: contig of 1630 bp in length
* 8300 8399: gap of unknown length
* 8400 9407: contig of 1008 bp in length
* 9408 9507: gap of unknown length
* 9508 10597: contig of 1090 bp in length
* 10598 10697: gap of unknown length
* 10698 11890: contig of 1193 bp in length
* 11891 11990: gap of unknown length
* 11991 13008: contig of 1018 bp in length
* 13009 13108: gap of unknown length
* 13109 15050: contig of 1942 bp in length
* 15051 15150: gap of unknown length
* 15151 16489: contig of 1339 bp in length
* 16490 16589: gap of unknown length
* 16590 17661: contig of 1072 bp in length
* 17662 17762: gap of unknown length
* 17762 19205: contig of 1444 bp in length
* 19206 19305: gap of unknown length
* 19306 20938: contig of 1633 bp in length
* 20939 21039: gap of unknown length
* 21039 22434: contig of 1396 bp in length
* 22435 22534: gap of unknown length
* 22535 23765: contig of 1231 bp in length
* 23766 23865: gap of unknown length
* 23866 25250: contig of 1385 bp in length
* 25251 25350: gap of unknown length
* 25351 26862: contig of 1512 bp in length
* 26863 26962: gap of unknown length
* 26963 28288: contig of 1326 bp in length
* 28289 28388: gap of unknown length

* 28389 30077: contig of 1689 bp in length
* 30078 30177: gap of unknown length
* 30178 31942: contig of 1765 bp in length
* 31943 32042: gap of unknown length
* 32043 32043: contig of 1001 bp in length
* 32044 33143: gap of unknown length
* 33144 35029: contig of 1886 bp in length
* 35030 35129: gap of unknown length
* 35130 36799: contig of 1670 bp in length
* 36800 36899: gap of unknown length
* 36900 39075: contig of 2176 bp in length
* 39076 39175: gap of unknown length
* 39176 40319: contig of 1144 bp in length
* 40320 40419: gap of unknown length
* 40420 42225: contig of 1806 bp in length
* 42226 42325: gap of unknown length
* 42326 44060: contig of 1735 bp in length
* 44061 44160: gap of unknown length
* 44161 46193: contig of 2033 bp in length
* 46194 46293: gap of unknown length
* 46294 47702: contig of 1409 bp in length
* 47703 47802: gap of unknown length
* 47803 49278: contig of 1476 bp in length
* 49279 49378: gap of unknown length
* 49379 50695: contig of 1317 bp in length
* 50696 50795: gap of unknown length
* 50796 52732: contig of 1937 bp in length
* 52733 52832: gap of unknown length
* 52833 54898: contig of 2066 bp in length
* 54899 54998: gap of unknown length
* 54999 58772: contig of 3774 bp in length
* 58773 58872: gap of unknown length
* 58873 60925: contig of 2053 bp in length
* 60926 61026: gap of unknown length
* 61026 63016: contig of 1991 bp in length
* 63017 63116: gap of unknown length
* 63117 66294: contig of 3178 bp in length
* 66295 66394: gap of unknown length
* 66395 69544: contig of 3150 bp in length
* 69545 69644: gap of unknown length
* 69645 72104: contig of 2460 bp in length
* 72105 72204: gap of unknown length
* 72205 74212: contig of 2008 bp in length
* 74213 74312: gap of unknown length
* 74313 76225: contig of 1913 bp in length
* 76226 76326: gap of unknown length
* 76326 79130: contig of 2805 bp in length
* 79131 79230: gap of unknown length
* 79231 81946: contig of 2716 bp in length
* 81947 82046: gap of unknown length
* 82047 86126: contig of 4080 bp in length
* 86127 86226: gap of unknown length
* 86227 91271: contig of 5045 bp in length
* 91272 91371: gap of unknown length
* 91372 93434: contig of 2063 bp in length
* 93435 100025: contig of 6491 bp in length
* 100026 100125: gap of unknown length
* 100126 103179: contig of 3054 bp in length
* 103180 103279: gap of unknown length
* 103280 106939: contig of 3660 bp in length
* 106940 107039: gap of unknown length
* 107040 111024: contig of 3985 bp in length
* 111025 111124: gap of unknown length
* 111125 117137: contig of 6013 bp in length
* 117138 117237: gap of unknown length
* 117238 121474: contig of 4237 bp in length
* 121475 121575: gap of unknown length
* 121575 126227: contig of 4653 bp in length
* 126228 126327: gap of unknown length
* 126328 132054: contig of 5727 bp in length
* 132055 132154: gap of unknown length
* 132155 138286: contig of 6132 bp in length

* 138287 138386: gap of unknown length																														
Query Match		20.6%	Score 352;	DB 2;	Length 155250;																									
Best Local Similarity		90.4%;	Pred. No. 6e-89;																											
Matches 376;		Conservative 0;	Mismatches 40;	Indels 0;	Gaps 0;																									
QY	344	AATCTGCAAGATATCTGGGAGAGATACCATATATCATCAACCCAGACAGAGTTCC 403																												
DB	81220	AAAATGCTGAGTACTGGGAGAGATCCCATCATCATCAGACAGACCAACAGAGTTCC 81161																												
QY	404	TTTCGATTCTCCCAACGGAGTTCCGCTGCTGGTGAAGTCCATGACCCACCCCTGCACCAA 463																												
DB	81160	TTTGATCTGCTGCCCTGGGAGTTCCGCTGCTGGTGAAGTCCATGACCCACCCCTGCACCA 81101																												
QY	464	CCTCAGCAACAACGCGAAGCCGCCCTACTATGCTGGACATCCCTCAGAGCCCATGTAGT 523																												
DB	81100	CCTCAGCAACAACGCGAAGCCGCCCTACTATGCTGGACATCCCTCAGAGCCCATGTAGT 81041																												
QY	524	CTCACCATCCATCAGTTCAGTTGATTCAGACACCGACGCTCTTCGCAACCTTATTGCC 583																												
DB	81040	CTCACCATCCATCAGTTCAGTTGATTCAGACACCGACGCTCTTCGCAACCTTATTGCC 80981																												
QY	584	ACAGCTCAGCGCCCAATCAGCAGACAGACAGGTGTAAAACTGAAGAGTGAACCT 643																												
DB	80980	ACAGCTCAGCGCCCAATCAGCAGACAGACAGGTGTAAAACTGAAGAGTGAACCT 80921																												
QY	644	CTTCCTCTGCTGCGGTGTCACCTCTCCTCCTGATGACCTTCGCTTTAGATGTPAAG 703																												
DB	80920	CTTCCTCTGCTGCGGTGTCACCTCTCCTCCTGATGACCTTCGCTTTAGATGTPAAG 80861																												
QY	704	AATCCCATGCACCATTCAGATCCGCGACAGTACCAGAGAGTGACTTTATCG 759																												
DB	80860	AATCCCATGCACCATTCAGATCCGCGACAGTACCAGAGAGTGACTTTATCG 80805																												
RESULT 13																														
AC096698/c																														
LOCUS	Rattus norvegicus clone CH230-154C12, *** SEQUENCING IN PROGRESS																													
DEFINITION	***, 68 unordered pieces.																													
ACCESSION	AC096698																													
VERSION	AC096698.7 GI:22218393																													
KEYWORDS	HTG; HTGS_PHASE1.																													
SOURCE	Norway rat.																													
ORGANISM	Rattus norvegicus																													
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;																														
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;																														
Rattus.																														
REFERENCE	1 (bases 1 to 145971)																													
AUTHORS	Muzny,D,Marie., Metzker,M,Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,O., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,K., Chen,X., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Haviak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,																													

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Louisedge, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Manquin, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okwuonu, G., Olarunpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, H., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 145971)
Worley, K.C.

Direct Submission
Submitted (23-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 145971)

REFERENCE
AUTHORS
JOURNAL

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Aug 14, 2002 this sequence version replaced gi:21735321.

COMMENT
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHMS
Center clone name: CH230-154C12
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 75503 bases at least Q40
Consensus quality: 79139 bases at least Q30
Consensus quality: 81960 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1518: contig of 1518 bp in length
1519: gap of unknown length
1618: contig of 1624 bp in length
1619: gap of unknown length
3242: contig of 1026 bp in length
3243: gap of unknown length
3343: contig of 1026 bp in length
3343: gap of unknown length
4368: contig of 1026 bp in length
4369: gap of unknown length
4469: contig of 1232 bp in length
5700: gap of unknown length
5701: contig of 1105 bp in length
5801: gap of unknown length
6905: contig of 1105 bp in length
6906: gap of unknown length
7006: contig of 1183 bp in length
8188: gap of unknown length
8189: contig of 1376 bp in length
8288: gap of unknown length
9664: contig of 1073 bp in length
9665: gap of unknown length
10837: contig of 1073 bp in length
10838: gap of unknown length
12099: contig of 1162 bp in length
12100: gap of unknown length
12100: contig of 1503 bp in length
13702: gap of unknown length
13703: contig of 1316 bp in length
13802: gap of unknown length
15118: contig of 1316 bp in length
15119: gap of unknown length
15218: contig of 1444 bp in length
15219: gap of unknown length
16662: contig of 1444 bp in length
16663: gap of unknown length
16762: contig of 1389 bp in length
18151: contig of 1389 bp in length
18152: gap of unknown length
20079: contig of 1828 bp in length
20080: gap of unknown length
20179: contig of 1272 bp in length
21451: contig of 1272 bp in length
21452: gap of unknown length
22704: contig of 1153 bp in length
22705: gap of unknown length
22804: contig of 1170 bp in length
23974: gap of unknown length
24074: contig of 1307 bp in length
25381: gap of unknown length
25382: contig of 1650 bp in length
25481: gap of unknown length
27131: contig of 1650 bp in length
27132: gap of unknown length
27232: contig of 1783 bp in length
29014: gap of unknown length
29114: contig of 1612 bp in length
29115: gap of unknown length
30726: contig of 1612 bp in length
30727: gap of unknown length
30826: contig of 1351 bp in length
32177: gap of unknown length
32178: contig of 1092 bp in length
33369: gap of unknown length
33469: contig of 1534 bp in length
35003: gap of unknown length
35103: contig of 1580 bp in length
36683: gap of unknown length
36783: contig of 1372 bp in length
38155: gap of unknown length
38255: contig of 1021 bp in length
39276: gap of unknown length
39277: contig of 1871 bp in length
41247: gap of unknown length
41248: contig of 1334 bp in length
42681: gap of unknown length
42781: contig of 1867 bp in length
44648: gap of unknown length
44748: contig of 1831 bp in length
46579: gap of unknown length
46580: contig of 1000 bp in length
46680: gap of unknown length
47779: contig of 1839 bp in length
49618: gap of unknown length
51952: contig of 2234 bp in length
51953: gap of unknown length

```
* 52053 53921: contig of 1869 bp in length
* 53922 54021: gap of unknown length
* 54022 55669: contig of 1648 bp in length
* 55669 55769: gap of unknown length
* 55769 57073: contig of 1304 bp in length
* 57073 57173: gap of unknown length
* 57173 59800: contig of 2627 bp in length
* 59800 61117: gap of unknown length
* 61117 61217: contig of 1217 bp in length
* 61217 62821: gap of unknown length
* 62821 62921: contig of 1604 bp in length
* 62921 65273: gap of unknown length
* 65273 65374: contig of 2352 bp in length
* 65374 68170: gap of unknown length
* 68170 68270: contig of 2797 bp in length
* 68270 69664: gap of unknown length
* 69664 69764: contig of 1394 bp in length
* 69764 70233: gap of unknown length
* 70233 72124: contig of 2259 bp in length
* 72124 74541: gap of unknown length
* 74541 76714: contig of 2418 bp in length
* 76714 76814: gap of unknown length
* 76814 79279: gap of unknown length
* 79279 81931: contig of 2465 bp in length
* 81931 82031: gap of unknown length
* 82031 84469: gap of unknown length
* 84469 84569: contig of 2552 bp in length
* 84569 86972: gap of unknown length
* 86972 86972: contig of 2438 bp in length
* 86972 86972: contig of 2403 bp in length

Query Match      14.8%; Score 253.2; DB 2: Length 145971;
Best Local Similarity 85.3%; Pred. No. 1.3e-60;
Matches 295; Conservative 0; Mismatches 48; Indels 3; Gaps 1;

QY 1303 AGCTTACCACGAGCTTCAAGTCAGAGGTAATGCTCTCTCTCTTTGGAACTCTGGCCCA 1362
Db 93957 ATCTTCCCTCCGTTTCTAGCTGAGAGGTAATGCTCTCTCTCTCTGGAACCTGGCTCA 94016

QY 1363 TGTGAATATGAGCTTCAAGAAAGTGRAGAGGCAATGCTCTCTGGGCAATGTGTGCTGGG 1422
Db 94017 TGTGAATATGAGCTTCAAGAAAGTGRAGAGGCAATGCTCTGCAATGCGCTGGTGGG 94076

QY 1423 CAGTGATCTCTTCAGAGAGCTATGTCTCAGGATGAGTGAAGCTGGGATTCCTCAGAGCCC 1482
Db 94077 CAGTGATCTCTTCAGGACCAATGTCTCAGGATGAGTGAAGCTGGGATTCCTCAGAGCCC 94136

QY 1483 TGATGACTCAGATAGCAGCTATGTTCCCACTCCACTGACAGCTCATGGGGTCTCTCCC 1542
Db 94137 TGAGGACTCAGACAGCAGCTATGTTCCCACTCCACTGATAGCTCATGGGGTCTCTCCC 94196

QY 1543 TGTTTTCACACGCGCTCAGAGAGAGGATGAGGAAATATAAAGAAAGAGGAGAT 1602
Db 94197 TGTTTTCACGCGCTCAGAGAGAGGATGAGGAAATATAAAGAAAGAGAT 94253

QY 1603 GTTTTTCAGACCTACTAGACCCCAACAGAAAGGTTTGTATTA 1648
Db 94254 TTTCAATCCAGACCTACAGAGCCAGCAGAAAGCCTTGATGTTAA 94299
```

```
RESULT 15
HS33F3/c
LOCUS HS33F3F 296 bp DNA linear PRI 18-OCT-1995
DEFINITION H.sapiens CpG island DNA genomic MseI fragment, clone 33f3, forward
read cp933f3.ftie.
ACCESSION Z58264
VERSION Z58264.1 GI:1029495
KEYWORDS CpG island; genomic MseI fragment.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE 1 (bases 1 to 296)
AUTHORS MacDonald,M., Huckle,E., Wilkinson,P. and Micklem,G.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
REFERENCE 2 (bases 1 to 296)
AUTHORS Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
TITLE Purification of CpG islands using a methylated DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
PUBMED 8012384
COMMENT Vector: pGEM-5zf(-)
Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hmp.mrc.ac.uk/ for details
or contact: biohelp@hmp.mrc.ac.uk.
FEATURES
source
location/Qualifiers
1..296
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="33f3"
/sex="male"
/tissue_type="blood"
/clone_lib="CGI-1"
/dev_stage="adult"
BASE COUNT 82 a 78 c 84 g 50 t 2 others
ORIGIN
Query Match 12.5%; Score 214.6; DB 9: Length 296;
Best Local Similarity 97.4%; Pred. No. 5.6e-50;
Matches 228; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
QY 27 AGGTCCAGACAGCGCCAGCCCGCTCTGCGACGACAGTGAATAGTGTGTACTCCT 86
Db 296 AGGTCCAGACAGCGCCAGCCCGCTCTGCGACGACAGTGAATAGTGTGTACTCCT 237

QY 87 TGTCTCGGTTTCAGGTCCAGACCTCCCGCTCTTCCGGCTGCCCTGAACGTCAGGGGACCTC 146
Db 236 TGTCTCGGTTTCAGGTCCAGACCTCCCGCTCTTCCGGCTGCCCTGAACGTCAGGGGACCTC 177

QY 147 AGGACCTGTGATTGGGCGCTTGGCGCGGCGGACCGTGCACGAGAAACCCCTGGAGGAC 206
Db 176 AGGACCTGTGATTGGGCGCTTGGCGCGGCGGACCGTGCACGAGAAACCCCTGGAGGAC 117

QY 207 TTGGGCATTCTTGGGCTCCGCTGCTTTTTCGTCGCTCTTTCGGGCAAGTA 259
Db 116 TTGGGCATTCTTGGGCTCCGCTGCTTTCGTCGCTCTTTCGGGCAAGTA 63
```

Search completed: July 18, 2003, 09:03:50
Job time : 4501 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:03:57 ; Search time 0.492891 Seconds
(without alignments)
673.193 Million cell updates/sec

Title: US-09-857-308-5
Perfect score: 37
Sequence: 1 LYQAVATI 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	33	89.2	413	1	NORB_VIBAL	Q56587	vibrio algi
2	30	81.1	312	1	LPXK_HELPJ	Q9zmb1	helicobacte
3	30	81.1	410	1	NORB_NEIMA	Q9jvp9	neisseria m
4	30	81.1	410	1	NORB_NEIMA	Q9k0m4	neisseria m
5	30	81.1	500	1	GAG_HV1RH	P05890	human immun
6	29	78.4	421	1	GLTT_BACCA	P24944	bacillus ca
7	29	78.4	421	1	GLTT_BACST	P24943	bacillus st
8	29	78.4	492	1	EAAT_ONCVO	Q25605	onchoerca
9	29	78.4	502	1	EA4A_CAEEL	Q22682	caenorhabdi
10	29	78.4	503	1	EA4A_CAEEL	Q10901	caenorhabdi
11	29	78.4	532	1	EA4A_CAEEL	Q21353	caenorhabdi
12	29	78.4	560	1	EA4A_HUMAN	O00341	homo sapien
13	29	78.4	572	1	EA4A_MOUSE	P43006	mus musculu
14	29	78.4	573	1	EA4A_RAT	P31596	rattus norv
15	29	78.4	574	1	EA4A_HUMAN	P43004	homo sapien
16	29	78.4	575	1	EA4A_CAEEL	Q21751	caenorhabdi
17	29	78.4	760	1	SM4A_MOUSE	O62178	mus musculu
18	29	78.4	1197	1	EVGS_ECO57	P58402	escherichia
19	29	78.4	1197	1	EVGS_ECOLI	P30855	escherichia
20	29	78.4	1226	1	YCS3_YEAST	P25357	saccharomyc
21	28	75.7	199	1	AROK_MYCLE	Q9ccs5	mycobacteri
22	28	75.7	233	1	GLP1_MESCR	P45852	mesembryant
23	28	75.7	312	1	LPXK_HELPJ	O25095	helicobacte
24	28	75.7	375	1	AMSC_ERWAM	Q46633	erwinia amy
25	28	75.7	413	1	NORB_VIBHA	Q9rfw0	vibrio harv
26	28	75.7	415	1	NORB_VIBCH	Q9kps2	vibrio chol
27	28	75.7	437	1	GLTP_ECOLI	P21345	escherichia
28	28	75.7	453	1	UCR2_BOVIN	P23004	bos taurus
29	28	75.7	462	1	CTXA_CHIOU	P58762	chiropsalmu
30	28	75.7	469	1	NIFE_SYNPH	O07355	synecococc
31	28	75.7	523	1	EA4A_MOUSE	P51906	mus musculu
32	28	75.7	523	1	EA4A_RAT	P51907	rattus norv
33	28	75.7	524	1	EA4A_BOVIN	Q95135	bos taurus

RESULT 1

ID	NORB_VIBAL	STANDARD;	PRT;	413 AA.
AC	Q56587;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Na(+)-translocating NADH-quinone reductase subunit B (EC 1.6.5.-)			
DE	(Na(+)-translocating NQR subunit B) (Na(+)-NQR subunit B) (NQR complex subunit B) (NQR-1 subunit B).			
GN	NQRB OR NQR2.			
OS	Vibrio alginolyticus.			
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=663;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NCIMB 11038;			
RX	MEDLINE=95104445; PubMed=7805867;			
RA	Beattie P., Tan K., Bourne R.M., Leach D.R.F., Rich P.R., Ward F.B.;			
RT	"Cloning and sequencing of four structural genes for the Na(+)-translocating NADH-ubiquinone oxidoreductase of Vibrio alginolyticus."			
RT	FEBS Lett. 356:333-338(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Hayashi M., Unemoto T., Sugiyama A.;			
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1-10.			
RX	MEDLINE=98149659; PubMed=9490015;			
RA	Nakayama Y., Hayashi M., Unemoto T.;			
RT	"Identification of six subunits constituting Na+-translocating NADH-quinone reductase from the marine Vibrio alginolyticus."			
RL	FEBS Lett. 422:240-242(1998).			
RN	[4]			
RP	SEQUENCE OF 1-6 AND 226-237, AND COFACTOR.			
RX	MEDLINE=20298664; PubMed=10838078;			
RA	Nakayama Y., Yasui M., Sugahara K., Hayashi M., Unemoto T.;			
RT	"Covalently bound flavin in the NqrB and NqrC subunits of Na(+)-translocating NADH-quinone reductase from Vibrio alginolyticus."			
RL	FEBS Lett. 474:165-168(2000).			
RN	[5]			
RP	INHIBITION OF ENZYMIC ACTIVITY.			
RX	MEDLINE=20016049; PubMed=10549856;			
RA	Nakayama Y., Hayashi M., Yoshikawa K., Mochida K., Unemoto T.;			
RT	"Inhibitor studies of a new antibiotic, koromicin, 2-n-heptyl-4-hydroxyquinoline N-oxide and Ag+ toward the Na+-translocating NADH-quinone reductase from the marine Vibrio alginolyticus."			
RL	Biol. Pharm. Bull. 22:1064-1067(1999).			
RN	[6]			
RP	COFACTOR, AND MASS SPECTROMETRY.			
RX	MEDLINE=21099804; PubMed=11163785;			
RA	Hayashi M., Nakayama Y., Yasui M., Maeda M., Furuishi K., Unemoto T.;			
RT	"FMN is covalently attached to a threonine residue in the NqrB and NqrC subunits of Na(+)-translocating NADH-quinone reductase from Vibrio alginolyticus."			
RT				

P43005 homo sapien
P31597 oryctolagus
Q04975 salmoneilla
P10563 emericella
P03161 ground squi
P97603 rattus norv
Q90610 gallus gall
P23615 saccharomyc
Q92859 homo sapien
P97798 mus musculu
Q65802 ebola virus
Q90683 gallus gall

RL FEBS Lett. 488:5-8(2001).
 RN [7]
 RP REVIEW.
 RX MEDLINE=21145117; PubMed=11248187;
 RA Hayashi M., Nakayama Y., Umemoto T.;
 RT "Recent progress in the Na(+)-translocating NADH-quinone reductase
 from the marine Vibrio alginolyticus.";
 RL Biochim. Biophys. Acta 1505:37-44(2001).
 RN [8]
 RP REVIEW.
 RX MEDLINE=21145118; PubMed=11248188;
 RA Steuber J.;
 RT "Na(+)-translocation by bacterial NADH:quinone oxidoreductases: an
 extension to the complex-I family of primary redox pumps.";
 RL Biochim. Biophys. Acta 1505:45-56(2001).
 CC -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
 UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
 OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
 ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
 UBISEMIQUINONE TO UBIQUINOL.
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone + Na(+)(In) = NAD(+) +
 ubiquinol + Na(+)(Out).
 CC -!- COFACTOR: FMN.
 CC -!- ENZYME REGULATION: THIS REACTION IS TIGHTLY COUPLED TO THE NA(+)-
 PUMPING ACTIVITY AND SPECIFICALLY REQUIRES NA(+) FOR ACTIVITY.
 CC INHIBITED BY KORMICIN AND 2-N-HEPTYL-4-HYDROXYQUINOLINE N-OXIDE
 (HONO).
 CC -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
 AND NQRF.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 (potential).
 CC -!- SIMILARITY: BELONGS TO THE NQRB/RNFD FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL; Z37111; CA85477.1; ALT_INIT.
 DR EMBL; AB080830; BAA22911.1; -.
 DR InterPro; IPR004338; NQR2_Rnfd_RnFe.
 DR Pfam; PF03116; NQR2_Rnfd_RnFe; 1.
 KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
 FT Flavoprotein; FMN; Transmembrane; Inner membrane.
 FT INIT_MET 0 0
 FT TRANSMEM 22 39 POTENTIAL.
 FT TRANSMEM 59 81 POTENTIAL.
 FT TRANSMEM 124 146 POTENTIAL.
 FT TRANSMEM 158 180 POTENTIAL.
 FT TRANSMEM 267 289 POTENTIAL.
 FT TRANSMEM 296 315 POTENTIAL.
 FT TRANSMEM 325 344 POTENTIAL.
 FT TRANSMEM 357 374 POTENTIAL.
 FT TRANSMEM 378 400 POTENTIAL.
 FT TRANSMEM 235 235 FMN.
 FT BINDING 228 228 A -> V (IN REF. 1).
 FT CONFLICT 228 228 W -> G (IN REF. 1).
 FT CONFLICT 294 294
 SQ SEQUENCE 413 AA; 45211 MW; A6F849F7A5C91C9D CRC64;
 Query Match 89.2%; Score 33; DB 1; Length 413;
 Best Local Similarity 75.0%; Pred. No. 4.2;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LYQAVATI 8
 Db 25 LYEAVATV 32

RESULT 2
 LPXK_HELPJ

ID LPXK_HELPJ STANDARD; PRT; 312 AA.
 AC Q9ZMB1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tetraacyldisaccharide 4'-kinase (EC 2.7.1.130) (Lipid A 4'-kinase).
 GN LPXK OR JHP0311.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 CC Helicobacter.
 CC NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nichelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 gastric pathogen Helicobacter pylori.";
 RL Nature 397:176-180(1999).
 CC -!- FUNCTION: TRANSFERS THE GAMMA-PHOSPHATE OF ATP TO THE 4'-POSITION
 OF A TETRAACYLDISACCHARIDE 1-PHOSPHATE INTERMEDIATE (TERMED DS-1-
 P) TO FORM TETRAACYLDISACCHARIDE 1,4'-BIS-PHOSPHATE (LIPID IVA)
 (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + 2,3-bis(3-hydroxytetradecanoyl)-D-
 glucosaminyl-(beta-D-1,6)-2,3-bis(3-hydroxytetradecanoyl)-D-
 glucosaminyl beta-phosphate = ADP + 2,3,2',3'-tetrakis(3-
 hydroxytetradecanoyl)-D-glucosaminyl-1,6-beta-D-glucosamine 1,4'-
 bisphosphate.
 CC -!- PATHWAY: Lipid A biosynthesis; sixth step.
 CC -!- SIMILARITY: BELONGS TO THE LPXK FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL; AE001467; AAD05884.1; -.
 DR InterPro; IPR003758; LpxK.
 DR Pfam; PF02606; LpxK; 1.
 DR TIGRFAMS; TIGR00682; lpxk; 1.
 KW Transferase; Kinase; Lipid A biosynthesis; Lipid synthesis;
 KW ATP-binding; Complete proteome.
 FT NP_BIND 60 67 ATP (POTENTIAL).
 SQ SEQUENCE 312 AA; 35567 MW; 43E44608F3A60FAA CRC64;
 Query Match 81.1%; Score 30; DB 1; Length 312;
 Best Local Similarity 62.5%; Pred. No. 16;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LYQAVATI 8
 Db 32 IYQCIATI 39
 RESULT 3
 NQRB_NEIMA STANDARD; PRT; 410 AA.
 AC Q9JVP9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Na(+)-translocating NADH-quinone reductase subunit B (EC 1.6.5.-)
 DE (Na(+)-translocating NQR subunit B) (Na(+)-NQR subunit B) (NQR complex
 subunit B) (NQR-1 subunit B).
 GN NQRB OR NMA0751.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

```
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature 404:502-506(2000).
CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone + Na(+)(In) = NAD(+) +
CC ubiquinol + Na(+)(Out).
CC -1- COFACTOR: FMN (BY SIMILARITY).
CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NORB, NQRC, NQRE
CC AND NQRF (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE NORB/RNFD FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL162754; CAB84034.1; -
DR InterPro; IPR004338; NQR2_Rnfd_Rnfe.
DR Pfam; PF03116; NQR2_Rnfd_Rnfe; 1.
DR Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
DR Flavoprotein; FMN; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 58 80 POTENTIAL.
FT TRANSMEM 120 142 POTENTIAL.
FT TRANSMEM 155 177 POTENTIAL.
FT TRANSMEM 264 286 POTENTIAL.
FT TRANSMEM 293 312 POTENTIAL.
FT TRANSMEM 322 341 POTENTIAL.
FT TRANSMEM 354 371 POTENTIAL.
FT TRANSMEM 375 397 POTENTIAL.
FT TRANSMEM 232 232 FMN (BY SIMILARITY).
FT BINDING 232 232
SQ SEQUENCE 410 AA; 44536 MW; D46528F211513ECE CRC64;

Query Match 81.1%; Score 30; DB 1; Length 410;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 26 LYEAATI 33

RESULT 4
NORB_NEIMB STANDARD; PRT; 410 AA.
AC Q9K0M4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit B (EC 1.6.5.-)
DE (Na(+)-translocating NQR subunit B) (Na(+)-NQR subunit B) (NQR complex
DE subunit B) (NQR-1 subunit B).
GN NORB OR NMB0568.
OS Neisseria meningitidis (serogroup B).
```

```
CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masiangni V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone + Na(+)(In) = NAD(+) +
CC ubiquinol + Na(+)(Out).
CC -1- COFACTOR: FMN (BY SIMILARITY).
CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NORB, NQRC, NQRE
CC AND NQRF (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE NORB/RNFD FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002412; AAF40996.1; -
DR TIGR; NMB0568; -
DR InterPro; IPR004338; NQR2_Rnfd_Rnfe.
DR Pfam; PF03116; NQR2_Rnfd_Rnfe; 1.
DR Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
DR Flavoprotein; FMN; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 58 80 POTENTIAL.
FT TRANSMEM 120 142 POTENTIAL.
FT TRANSMEM 155 177 POTENTIAL.
FT TRANSMEM 264 286 POTENTIAL.
FT TRANSMEM 293 312 POTENTIAL.
FT TRANSMEM 322 341 POTENTIAL.
FT TRANSMEM 354 371 POTENTIAL.
FT TRANSMEM 375 397 POTENTIAL.
FT TRANSMEM 232 232 FMN (BY SIMILARITY).
FT BINDING 232 232
SQ SEQUENCE 410 AA; 44601 MW; F632E12206170B4F CRC64;

Query Match 81.1%; Score 30; DB 1; Length 410;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 26 LYEAATI 33

RESULT 5
GAG_HV1RH STANDARD; PRT; 500 AA.
AC P05890;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GAG polyprotein [Contains: Core proteins P17, P24, P2, P7, P1, P6].
```

GN GAG.
OS Human immunodeficiency virus type 1 (RF/HAT isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11701;
RN [1]
RP SEQUENCE FROM N.A.
RA Starcich B.R., Hahn B.H., Shaw G.M., McNeely P.D., Modrow S.,
RA Wolf H., Parks E.S., Parks W.P., Josephs S.F., Gallo R.C.,
RA Wong-Staal F.;
RL Submitted (XXX-1987) to the HIV data bank.
CC -!- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE
CC ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL
CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM
CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY
CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.
CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY
CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.
CC -!- PTM: THE P24 PROTEIN IS PHOSPHORYLATED.
CC -!- SIMILARITY: CONTAINS 2 CCHC-TYPE ZINC FINGERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M17451; AAA45052.1; -
CC HSSP: P05888; IAAF.
CC HIV: M17451; GAG\$RF.
CC InterPro: IPR000721; Gag_p24.
CC InterPro: IPR000071; Retrovir_p17.
CC InterPro: IPR001878; Znf_CCHC.
CC Pfam: PF00098; zf-CCHC; 2.
CC Pfam: PF00540; Gag_p17; 1.
CC Pfam: PF00607; Gag_p24; 1.
CC PRINTS: PR00939; C2HCZNFINGER.
CC PRINTS: PR00234; HIVIMATRIX.
CC SMART: SM00343; Znf_C2HC; 2.
CC PROSITE: PS0158; ZF_CCHC; 2.
KW AIDS: Core protein; Polyprotein; Myristate; Phosphorylation;
KW Zinc-finger; Repeat.
FT INIT_MET 0 BY SIMILARITY.
FT CHAIN 1 131 CORE PROTEIN P17 (MATRIX PROTEIN).
FT CHAIN 132 362 CORE PROTEIN P24 (CORE ANTIGEN).
FT CHAIN 363 376 CORE PROTEIN P2.
FT CHAIN 377 433 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).
FT CHAIN 434 449 CORE PROTEIN P1.
FT CHAIN 450 500 CORE PROTEIN P6.
FT ZN_FING 389 406 CCHC-TYPE 1.
FT ZN_FING 410 427 CCHC-TYPE 2.
FT LIPID 1 MYRISTATE (BY SIMILARITY).
SQ SEQUENCE 500 AA; 55825 MW; 44AA0CB5CD9EFF7B CRC64;

Query Match 81.1%; Score 30; DB 1; Length 500;
Best Local Similarity 75.0%; Pred. No. 26;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 77 LYNAVATL 84

RESULT 6
GLTT_BACCA
ID GLTT_BACCA STANDARD; PRT; 421 AA.
AC P24944;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proton/sodium-glutamate symport protein (Glutamate-aspartate carrier
DE protein).

GN GLTT.
OS Bacillus caldotenax.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1395;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93062018; PubMed=1359385;
RA Tolner B., Poolman B., Konings W.N.;
RT "Characterization and functional expression in Escherichia coli of
RT the sodium/proton/glutamate symport proteins of Bacillus
RT stearothermophilus and Bacillus caldotenax.";
RL Mol. Microbiol. 6:2845-2856(1992).
CC -!- FUNCTION: THIS CARRIER PROTEIN IS PART OF THE NA(+)-DEPENDENT,
CC BINDING-PROTEIN-INDEPENDENT GLUTAMATE-ASPARTATE TRANSPORT SYSTEM.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE SODIUM-DICARBOXYLATE SYMPORTER FAMILY
CC (SDF, TC 2.A.23).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M86509; AAA22493.1; -
CC PIR: S26246; S26246.
CC InterPro: IPR001991; Na/dico_symp.
CC Pfam: PF00375; SDF; 1.
CC PROSITE: PS00713; NA_DICARBOXYL_SYM_1; 1.
CC PROSITE: PS00714; NA_DICARBOXYL_SYM_2; 1.
KW Transport; Transmembrane; Symport.
FT DOMAIN 1 3 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 4 24 POTENTIAL.
FT DOMAIN 25 43 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 44 64 POTENTIAL.
FT DOMAIN 65 77 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 78 98 POTENTIAL.
FT DOMAIN 99 148 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 149 169 POTENTIAL.
FT DOMAIN 170 198 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 199 219 POTENTIAL.
FT DOMAIN 220 222 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 223 243 POTENTIAL.
FT DOMAIN 244 244 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 245 265 POTENTIAL.
FT DOMAIN 266 306 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 307 327 POTENTIAL.
FT DOMAIN 328 330 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 331 351 POTENTIAL.
FT TRANSMEM 352 372 POTENTIAL.
FT DOMAIN 373 421 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 421 AA; 45345 MW; 10ABBE12EDD1E7E4 CRC64;

Query Match 78.4%; Score 29; DB 1; Length 421;
Best Local Similarity 75.0%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 308 LYQALAAI 315

RESULT 7
GLTT_BACST
ID GLTT_BACST STANDARD; PRT; 421 AA.
AC P24943;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proton/sodium-glutamate symport protein (Glutamate-aspartate carrier
DE protein).

GN GLTT.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillales; Geobacillus.
 OX NCBI_TaxID=1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 7954;
 RX MEDLINE=93062018; PubMed=1359385;
 RA Tolner B., Poolman B., Konings W.N.;
 RT "Characterization and functional expression in Escherichia coli of
 the sodium/proton/glutamate symport proteins of Bacillus
 stearothermophilus and Bacillus caldotenax.";
 RL Mol. Microbiol. 6:2845-2856(1992).
 CC -!- FUNCTION: THIS CARRIER PROTEIN IS PART OF THE NA(+)-DEPENDENT,
 BINDING-PROTEIN-INDEPENDENT GLUTAMATE-ASPARTATE TRANSPORT SYSTEM.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
 (SDF, TC 2.A.23).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M86508; AAA22492.1; -;
 DR PIR; S26247; S26247.
 DR InterPro; IPR001991; Na/dico_sympt.
 DR Pfam; PF00375; SDF; 1.
 DR PROSITE; PS00713; NA_DICARBOXYL_SYMP_1; 1.
 DR PROSITE; PS00714; NA_DICARBOXYL_SYMP_2; 1.
 DR Transport; Transmembrane; Symport.
 FT DOMAIN 1 3 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 4 24 POTENTIAL.
 FT DOMAIN 25 43 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 44 64 POTENTIAL.
 FT DOMAIN 65 77 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 78 98 POTENTIAL.
 FT DOMAIN 99 148 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 149 169 POTENTIAL.
 FT DOMAIN 170 198 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 199 219 POTENTIAL.
 FT DOMAIN 220 222 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 223 243 POTENTIAL.
 FT DOMAIN 244 244 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 245 265 POTENTIAL.
 FT DOMAIN 266 306 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 307 327 POTENTIAL.
 FT DOMAIN 328 330 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 331 351 POTENTIAL.
 FT TRANSMEM 352 372 POTENTIAL.
 FT DOMAIN 373 421 CYTOPLASMIC (POTENTIAL).
 SQ SEQUENCE 421 AA; 45469 MW; 30C5738E8FD3A54F CRC64;
 Query Match 78.4%; Score 29; DB 1; Length 421;
 Best Local Similarity 75.0%; Pred. No. 37;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LYQAVATI 8
 Db 308 LYQALAAI 315
 RESULT 8
 ID EAAT_ONCVO STANDARD; PRT; 492 AA.
 AC Q25605;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Excitatory amino acid transporter (Sodium-dependent glutamate/

DE aspartate transporter).
 GN GLT-1.
 OS Onchocerca volvulus.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 OC Onchocercidae; Onchocerca.
 OX NCBI_TaxID=6282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97039669; PubMed=8885221;
 RA Radice A.D., Lustigman S.;
 RT "Cloning and characterization of cDNAs encoding putative glutamate
 transporters from Caenorhabditis elegans and Onchocerca volvulus.";
 RL Mol. Biochem. Parasitol. 80:41-53(1996).
 CC -!- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
 ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
 RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
 AS A SYMPORT BY CO-TRANSPORTING SODIUM (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
 (SDF, TC 2.A.23).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U35251; AAB41937.1; -;
 DR InterPro; IPR001991; Na/dico_sympt.
 DR Pfam; PF00375; SDF; 1.
 DR PRINTS; PR00173; EDTRNSPORT.
 DR PROSITE; PS00713; NA_DICARBOXYL_SYMP_1; 1.
 DR PROSITE; PS00714; NA_DICARBOXYL_SYMP_2; 1.
 DR Transport; Transmembrane; Glycopolymers; Symport.
 FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 8 28 POTENTIAL.
 FT TRANSMEM 47 67 POTENTIAL.
 FT TRANSMEM 85 105 POTENTIAL.
 FT DOMAIN 106 191 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 192 212 POTENTIAL.
 FT TRANSMEM 228 248 POTENTIAL.
 FT TRANSMEM 270 290 POTENTIAL.
 FT TRANSMEM 358 378 POTENTIAL.
 FT TRANSMEM 389 409 POTENTIAL.
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 492 AA; 53391 MW; CEE52D670F76A89E CRC64;
 Query Match 78.4%; Score 29; DB 1; Length 492;
 Best Local Similarity 75.0%; Pred. No. 43;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LYQAVATI 8
 Db 355 LYEAVALI 362
 RESULT 9
 ID EAAT_CAEEL STANDARD; PRT; 502 AA.
 AC Q22682;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative sodium-dependent excitatory amino acid transporter T22E5.2.
 GN T22E5.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Minx P.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC (SDF, TC 2.A.23).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U43282; AAA83614.1; -;
DR WormPep: T22E5.2; CE04992.
DR InterPro: IPR001991; Na/dico_symp.
DR Pfam: PF00375; SDF; 1.
DR PRINTS: PR00173; EDTNSPORT.
DR PROSITE: PS00713; NA_DICARBOXYL_SYMPT_1; 1.
DR PROSITE: PS00714; NA_DICARBOXYL_SYMPT_2; 1.
KW Hypothetical protein; Transport; transmembrane; Glycoprotein; Symport.
FT DOMAIN 1 31 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 32 52 POTENTIAL.
FT TRANSMEM 74 94 POTENTIAL.
FT TRANSMEM 111 131 POTENTIAL.
FT DOMAIN 132 217 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 218 238 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 300 320 POTENTIAL.
FT TRANSMEM 388 408 POTENTIAL.
FT TRANSMEM 419 439 POTENTIAL.
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 502 AA; 54610 MW; 8B4C7469C594727D CRC64;

Query Match 78.4%; Score 29; DB 1; Length 502;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
II:III I
Db 385 LYEAVAAI 392

RESULT 10
EAAI_CAEEL STANDARD; PRT: 503 AA.
AC Q10901; P90798; Q17920;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Excitatory amino acid transporter (Sodium-dependent glutamate/
DE aspartate transporter).
GN Glt-1 or c12d12.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RX MEDLINE=97039669; PubMed=8885221;
RA Radice A.D., Lustigman S.;
RT "Cloning and characterization of cDNAs encoding putative glutamate
RT transporters from Caenorhabditis elegans and Onchocerca volvulus.";
RL Mol. Biochem. Parasitol. 80:41-53(1996).
[2]
RN SEQUENCE FROM N.A. (SHORT FORM).
RC STRAIN-Bristol N2;
RX MEDLINE=97079197; PubMed=8920929;
RA Kawano T., Takuwa K., Nakajima T.;

RT "Molecular cloning of a cDNA for the glutamate transporter of the
RL nematode Caenorhabditis elegans.";
RL Biochem. Biophys. Res. Commun. 228:415-420(1996).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN-Bristol N2;
RA Nhan M., Hawkins J.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
[4]
RN REVISIONS.
RP Waterston R.;
RA Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A.
RP STRAIN-Bristol N2;
RC MEDLINE=97321871; PubMed=9178573;
RX Kawano T., Takuwa K., Nakajima T.;
RA "Structure and activity of a new form of the glutamate transporter of
RT the nematode Caenorhabditis elegans.";
RL Biosci. Biotechnol. Biochem. 61:927-929(1997).
CC -!- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
CC ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
CC RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
CC AS A SYMPORT BY CO-TRANSPORTING SODIUM (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; GLT-1 AND GLT-2 (SHOWN HERE);
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC (SDF, TC 2.A.23).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U35250; AAB41909.1; -;
DR EMBL: U35250; AAB41910.1; -;
DR EMBL: D86740; BAA13164.1; -;
DR EMBL: U51998; AAL00857.1; -;
DR EMBL: U51998; AAL00858.1; -;
DR EMBL: D86741; BAA21840.1; -;
DR WormPep: C12D12.2a; CE29083.
DR WormPep: C12D12.2b; CE29084.
DR InterPro: IPR001991; Na/dico_symp.
DR Pfam: PF00375; SDF; 1.
DR PRINTS: PR00173; EDTNSPORT.
DR PROSITE: PS00713; NA_DICARBOXYL_SYMPT_1; 1.
DR PROSITE: PS00714; NA_DICARBOXYL_SYMPT_2; 1.
KW Transport; Transmembrane; Glycoprotein; Symport; Alternative splicing.
FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 19 39 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 96 116 POTENTIAL.
FT DOMAIN 117 198 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 239 259 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 369 389 POTENTIAL.
FT TRANSMEM 400 420 POTENTIAL.
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 11 MISSING (IN ISOFORM GLT-1).
FT CONFLICT 68 68 L -> P (IN REF. 1).
FT CONFLICT 270 273 ADTA -> ENTT (IN REF. 1).
FT CONFLICT 321 321 A -> G (IN REF. 1).
FT CONFLICT 480 503 AMNDEKRLQAVYNSLPDDEKHTH -> GFVF (IN
FT REF. 3).
SQ SEQUENCE 503 AA; 54675 MW; 3014BFF28E41E798 CRC64;

Query Match 78.4%; Score 29; DB 1; Length 503;

Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 366 LYEAVAI 373

RESULT 11

EEA2_CAEEL STANDARD; PRT; 532 AA.
AC Q21353;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative sodium-dependent excitatory amino acid transporter K08F4.4.
GN K08F4.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Hendry C.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
(SDF, TC 2.A.23).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; Z68879; CAA93084.1; -;
DR WormPep; K08F4.4; CE06152.
DR InterPro; IPR001991; Na/dico_symp.
DR Pfam; PF00375; SDF; 1.
DR PRINTS; PR00173; EDRNSPORT.
DR PROSITE; PS00713; NA_DICARBOXYL_SYMPT_1; 1.
DR PROSITE; PS00714; NA_DICARBOXYL_SYMPT_2; 1.
KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport.
FT DOMAIN 1 5 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 46 66 POTENTIAL.
FT TRANSMEM 83 103 POTENTIAL.
FT DOMAIN 104 181 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 222 242 POTENTIAL.
FT TRANSMEM 264 284 POTENTIAL.
FT TRANSMEM 352 372 POTENTIAL.
FT TRANSMEM 383 402 POTENTIAL.
FT CARBOHYD 164 164 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 169 169 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 532 AA; 58469 MW; FA370EAD3795E5DF CRC64;

Query Match 78.4%; Score 29; DB 1; Length 532;
Best Local Similarity 75.0%; Pred. No. 47;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 349 LYEAVAI 356

RESULT 12

EEA5_HUMAN STANDARD; PRT; 560 AA.
ID EAA5_HUMAN
AC O00341;

DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Excitatory amino acid transporter 5 (Retinal glutamate transporter).
GN SLC1A7 OR EAAT5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=97268713; PubMed=9108121;
RA Arriza J.L., Eliasof S., Kavanaugh M.P., Amara S.G.;
RT "Excitatory amino acid transporter 5, a retinal glutamate transporter
coupled to a chloride conductance.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:4155-4160(1997).
CC -!- FUNCTION: TRANSPORTS L-GLUTAMATE; THE L-GLUTAMATE UPTAKE IS
SODIUM- AND VOLTAGE-DEPENDENT AND CHLORIDE-INDEPENDENT. THE EAAT5-
ASSOCIATED CHLORIDE CONDUCTANCE MAY PARTICIPATE IN VISUAL
PROCESSING.
CC -!- SUBUNIT: INTERACTS WITH THE PDZ DOMAINS OF DLG4.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN RETINA. DETECTABLE IN
LIVER, HEART, MUSCLE AND BRAIN.
CC -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
(SDF, TC 2.A.23).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U76362; AAB53971.1; -;
DR Genew; HGNC:10945; SLC1A7.
DR MIN; G04471; -;
DR InterPro; IPR001991; Na/dico_symp.
DR Pfam; PF00375; SDF; 1.
DR PRINTS; PR00173; EDRNSPORT.
DR PROSITE; PS00713; NA_DICARBOXYL_SYMPT_1; 1.
DR PROSITE; PS00714; NA_DICARBOXYL_SYMPT_2; 1.
KW Transport; Transmembrane; Glycoprotein; Symport; Multigene family.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 37 POTENTIAL.
FT TRANSMEM 60 80 POTENTIAL.
FT TRANSMEM 94 114 POTENTIAL.
FT DOMAIN 115 216 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 217 237 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 300 320 POTENTIAL.
FT TRANSMEM 330 350 POTENTIAL.
FT TRANSMEM 372 392 POTENTIAL.
FT TRANSMEM 414 434 POTENTIAL.
FT TRANSMEM 457 477 POTENTIAL.
FT CARBOHYD 191 191 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 560 AA; 60722 MW; C71D48355AACFF32 CRC64;

Query Match 78.4%; Score 29; DB 1; Length 560;
Best Local Similarity 75.0%; Pred. No. 50;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 383 LYEAVAI 390

RESULT 13

EAA2_MOUSE
ID EAA2_MOUSE STANDARD; PRT; 572 AA.
AC P43006; O35877; O54686; O54687;

DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Excitatory amino acid transporter 2 (Sodium-dependent
 DE glutamate/aspartate transporter 2) (GLT-1).
 GN SLCA12 OR EAAT2 OR GLT1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=95213010; PubMed=7698742;
 RA Kirschner M.A., Copeland N.G., Gilbert D.J., Jenkins N.A.,
 RA Anara S.G.;
 RT "Mouse excitatory amino acid transporter EAAT2: isolation,
 RT characterization, and proximity to neuroexcitability loci on mouse
 RT chromosome 2.";
 RL Genomics 24:218-224(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JCL:ICR; TISSUE=Cerebellum;
 RX MEDLINE=95284091; PubMed=766664;
 RA Mukainaka Y., Tanaka K., Hagiwara T., Wada K.;
 RT "Molecular cloning of two glutamate transporter subtypes from mouse
 RT brain.";
 RL Biochim. Biophys. Acta 1244:233-237(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=96032356; PubMed=7557442;
 RA Sutherland M.L., Delaney T.A., Noebels J.L.;
 RT "Molecular characterization of a high-affinity mouse glutamate
 RT transporter.";
 RL Gene 162:271-274(1995).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Peng J.-B., Guo L.-H.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC STRAIN=JCL:ICR; TISSUE=Brain, and Liver;
 RX MEDLINE=98039013; PubMed=9373176;
 RA Utsumomiya-Tate N., Endou H., Kanai Y.;
 RT "Tissue specific variants of glutamate transporter GLT-1.";
 RL FEBS Lett. 416:312-316(1997).
 CC -1- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
 CC ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
 CC RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
 CC AS A SYMPORT BY CO-TRANSPORTING SODIUM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; GLT-1 (SHOWN HERE), GLT-1A
 CC AND GLT-1B; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: ISOFORM GLT1 IS EXPRESSED IN THE BRAIN.
 CC ISOFORMS GLT-1A AND GLT-1B ARE EXPRESSED IN THE LIVER.
 CC -1- PTM: GLYCOSYLATED.
 CC -1- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
 CC (SDF, TC 2.A.23).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 DR EMBL; U11763; AAA77673.1; -;
 DR EMBL; D43796; BAA07854.1; -;
 DR EMBL; U24699; AAA91643.1; -;
 DR EMBL; U75372; AAB71737.1; -;

DR EMBL; U75373; AAB71738.1; -;
 DR EMBL; AB007810; BAA23770.1; -;
 DR EMBL; AB007811; BAA23771.1; -;
 DR EMBL; AB007812; BAA23772.1; -;
 DR MGD; MGI:101931; Slc1a2.
 DR InterPro: IPR001991; Na/dico_symp.
 DR Pfam: PF00375; SDF; 1.
 DR PRINTS: PR00173; EDTRNSPORT.
 DR PROSITE: PS00713; NA_DICARBOXYL_SYMP_1; 1.
 DR PROSITE: PS00714; NA_DICARBOXYL_SYMP_2; 1.
 KW Transport; transmembrane; Glycoprotein; Symport; Multigene family;
 KW Alternative splicing.
 FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 45 64 POTENTIAL.
 FT TRANSMEM 88 108 POTENTIAL.
 FT TRANSMEM 121 142 POTENTIAL.
 FT DOMAIN 143 238 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 239 258 POTENTIAL.
 FT TRANSMEM 279 300 POTENTIAL.
 FT TRANSMEM 316 338 POTENTIAL.
 FT TRANSMEM 405 429 POTENTIAL.
 FT TRANSMEM 436 458 POTENTIAL.
 FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1 6 MASTEG -> MVS (IN ISOFORMS GLT-1A AND
 FT ISOFORM GLT-1B).
 FT VARSPLIC 551 572 TLAANGKSADCSVEEPPKREK -> PPFPLDIETCI
 FT (IN ISOFORM GLT-1B).
 FT CONFLICT 26 26 D -> E (IN REF. 3).
 FT CONFLICT 62 62 G -> R (IN REF. 3).
 FT CONFLICT 112 112 A -> V (IN REF. 3).
 FT CONFLICT 454 454 T -> I (IN REF. 4).
 FT CONFLICT 525 525 K -> L (IN REF. 4).
 FT CONFLICT 572 572 K -> EFD (IN REF. 3).
 SQ SEQUENCE 572 AA; 62030 MW; 13C7C30DED40CA81 CRC64;
 Query Match 78.4%; Score 29; DB 1; Length 572;
 Best Local Similarity 75.0%; Pred. No. 51;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LYQAVATI 8
 Db 402 LYEAVAAI 409
 RESULT 14
 EAA2_RAT
 ID EAA2_RAT STANDARD; PRT; 573 AA.
 AC P31596;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Excitatory amino acid transporter 2 (Sodium-dependent
 DE glutamate/aspartate transporter 2) (GLT-R) (GLT-1).
 GN SLCA12 OR EAAT2 OR GLT1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93078876; PubMed=1448170;
 RA Pines G., Danbolt N.C., Bjoeraas M., Zhang Y., Bendahan A., Eide L.,
 RA Koepsell H., Storm-Mathisen J., Seeborg E., Kanner B.I.;
 RT "Cloning and expression of a rat brain L-glutamate transporter.";
 RL Nature 360:464-467(1992).
 RN [2]
 RP ERRATUM.
 RA Pines G., Danbolt N.C., Bjoeraas M., Zhang Y., Bendahan A., Eide L.,
 RA Koepsell H., Storm-Mathisen J., Seeborg E., Kanner B.I.;
 RL Nature 360:768-768(1992).
 RN [3]

RP REVISIONS TO 260-289.
RX MEDLINE-93292659; PubMed=8099882;
RA Kanner B.I.;
RT "Glutamate transporters from brain. A novel neurotransmitter
transporter family";
RL FEBS Lett. 325:95-99(1993).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Forebrain;
RA Roginski R.S., Choudhury K., Meiners S., Marone M., Basma A.N.,
RA Geller H.M.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM GLT-1A).
RC TISSUE=Hepatoma;
RX MEDLINE-20521659; PubMed=11068035;
RA Pollard M., McGivan J.;
RT "The rat hepatoma cell line H4-II-E-C3 expresses high activities of
the high-affinity glutamate transporter GLT-1A";
RL FEBS Lett. 484:74-76(2000).
RN [6]
RP MUTAGENESIS OF LYS-298 AND HIS-326.
RX MEDLINE-94308247; PubMed=7913472;
RA Zhang Y., Pines G., Kanner B.I.;
RT "Histidine 326 is critical for the function of GLT-1, a (Na⁺ + K⁺)-
coupled glutamate transporter from rat brain";
RL J. Biol. Chem. 269:19573-19577(1994).
CC -1- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
CC ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
CC RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
CC AS A SYMPORT BY CO-TRANSPORTING SODIUM.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; GLT1 (SHOWN HERE) AND GLT-1A;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: LOCALIZED IN BRAIN AND IS HIGHLY ENRICHED IN
CC THE PURKINJE CELL LAYER IN CEREBELLUM.
CC -1- PTM: GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC (SDF, TC 2.A.23).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X67857; CAM48042.1; ALT_SEQ.
DR EMBL; U15098; AAA93061.1; -
DR EMBL; U15098; AAA93062.1; ALT_INIT.
DR EMBL; AF297648; AAG13411.1; -
DR InterPro: IPR001991; Na/dico_symp.
DR Pfam: PF00375; SDF; 1.
DR PRINTS: PR00173; EDPRNSPORT.
DR PROSITE: PS00713; NA_DICARBOXYL_SYMPT_1; 1.
DR PROSITE: PS00714; NA_DICARBOXYL_SYMPT_2; 1.
DR TRANSPORT; Transmembrane; Glycoprotein; Symport; Multigene family;
KW Alternative splicing.
FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 45 64 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 121 142 POTENTIAL.
FT DOMAIN 143 238 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 239 258 POTENTIAL.
FT TRANSMEM 279 300 POTENTIAL.
FT TRANSMEM 316 338 POTENTIAL.
FT TRANSMEM 405 429 POTENTIAL.
FT TRANSMEM 436 458 POTENTIAL.
FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 1 6 MASTEP -> MVS (IN ISOFORM GLT-1A).
FT MUTAGEN 298 K->N,T; REDUCED TRANSPORTER ACTIVITY.

FT MUTAGEN 298 K->H,R; NORMAL TRANSPORTER ACTIVITY.
FT MUTAGEN 326 H->N,T,K,R; NO TRANSPORTER ACTIVITY.
FT CONFLICT 521 V -> I (IN REF. 4).
SQ SEQUENCE 573 AA; 62106 MW; 8C51D30954E00E7F CRC64;
Query Match 78.4%; Score 29; DB 1; Length 573;
Best Local Similarity 75.0%; Pred. No. 51;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LYQAVATI 8
DB 402 LYEAVALI 409
RESULT 15
EAA2_HUMAN STANDARD; PRT; 574 AA.
ID EAA2_HUMAN
AC P43004; Q14417;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Excitatory amino acid transporter 2 (Sodium-dependent
DE glutamate/aspartate transporter 2).
GN SLC1A2 OR EAAT2 OR GLT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain Cortex;
RX MEDLINE-94365697; PubMed=7521911;
RA Arriza J.L., Fairman W.A., Wendy A., Wadiche J.I., Murdoch G.H.,
RA Kavanaugh M.P., Amara S.G.;
RT "Functional comparisons of three glutamate transporter subtypes
RT cloned from human motor cortex";
RL J. Neurosci. 14:5559-5569(1994).
RN [2]
RP SEQUENCE OF 11-574 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE-94227088; PubMed=8172925;
RA Shashidharan P., Wittenberg I., Plaitakis A.;
RT "Molecular cloning of human brain glutamate/aspartate transporter
RT II";
RL Biochim. Biophys. Acta 1191:393-396(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE-95002073; PubMed=7522567;
RA Manfras B.J., Rudert W.A., Trucco M., Boehm B.O.;
RT "Cloning and characterization of a glutamate transporter cDNA from
RT human brain and pancreas";
RL Biochim. Biophys. Acta 1195:185-188(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE-98130689; PubMed=9463476;
RA Shimamoto K., Lebrun B., Yasuda-Kamatani Y., Sakaitani M., Shigeri Y.,
RA Yumoto N., Nakajima T.;
RT "Di-thereo-b-benzoyloxyaspartate, a potent blocker of excitatory amino
RT acid transporters";
RL Mol. Pharmacol. 53:195-201(1998).
CC -1- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
CC ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
CC RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
CC AS A SYMPORT BY CO-TRANSPORTING SODIUM.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- PTM: GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC (SDF, TC 2.A.23).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U03505; AAA50429.1; -
DR EMBL; U01824; AAA18900.1; -
DR EMBL; Z32517; CAA83532.1; -
DR EMBL; D85884; BAA28706.1; -
DR Genew; HGNC:10940; SLC1A2.
DR MIM; 600300; -
DR InterPro; IPR001991; Na/dico_symp.
DR Pfam; PF00375; SDF; 1.
DR PRINTS; PR00173; EDTRNSPORT.
DR PROSITE; PS00713; NA_DICARBOXYL_SYMP.1; 1.
DR PROSITE; PS00714; NA_DICARBOXYL_SYMP.2; 1.
KW Transport; Transmembrane; Glycoprotein; Symport; Multigene family.
FT DOMAIN 1 44
FT TRANSMEM 45 64
FT TRANSMEM 88 108
FT TRANSMEM 121 142
FT TRANSMEM 143 239
FT DOMAIN 240 259
FT TRANSMEM 280 301
FT TRANSMEM 317 339
FT TRANSMEM 406 430
FT TRANSMEM 437 459
FT CARBOHYD 206 206
FT CONFLICT 216 216
FT CONFLICT 19 19
FT CONFLICT 27 27
FT CONFLICT 50 50
FT CONFLICT 58 58
FT CONFLICT 141 141
FT CONFLICT 155 155
FT CONFLICT 211 211
FT CONFLICT 253 255
FT CONFLICT 263 289
FT CONFLICT 347 347
FT CONFLICT 539 539
FT CONFLICT 556 556
FT CONFLICT 563 565
FT CONFLICT 570 570
SQ SEQUENCE 574 AA; 62104 MW; C8104B6727979435 CRC64;
PERDQNEVSDHDDHVL (IN REF. 3).
F -> L (IN REF. 1).
Y -> F (IN REF. 3).
A -> G (IN REF. 3).
CSV -> RVL (IN REF. 3).
W -> G (IN REF. 3).
GIA -> AIP (IN REF. 3).
AKLMVDFENILNEIVMKLVIMIMWYSP -> GQADGGFLQH
N-LINKED (GLCNAC. .) (POTENTIAL).
H -> P (IN REF. 1).
E -> G (IN REF. 3).
T -> Q (IN REF. 2).
A -> S (IN REF. 2).
A -> G (IN REF. 3).
P -> A (IN REF. 3).
V -> E (IN REF. 1).
PERDQNEVSDHDDHVL (IN REF. 3).

Query Match 78.4%; Score 29; DB 1; Length 574;
Best Local Similarity 75.0%; Pred. No. 51;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LYQAVATI 8
DB 403 LYEAVAAI 410

Search completed: July 18, 2003, 09:57:52
Job time : 2.49289 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:55:28 ; Search time 0.606635 Seconds
(without alignments)
388.015 Million cell updates/sec

Title: US-09-857-308-5
Perfect score: 37
Sequence: 1 LYQAVATI 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_A1:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	37	100.0	363	4	US-09-461-697-81
2	37	100.0	412	4	US-09-461-697-77
3	30	81.1	20	3	US-08-695-301A-30
4	30	81.1	20	3	US-08-695-304C-30
5	30	81.1	20	4	US-09-589-768-30
6	30	81.1	20	4	US-09-594-845-30
7	29	78.4	560	2	US-08-948-569A-10
8	29	78.4	560	2	US-09-188-469-10
9	29	78.4	560	4	US-09-397-238A-10
10	29	78.4	564	2	US-08-948-569A-12
11	29	78.4	564	2	US-09-188-469-12
12	29	78.4	564	4	US-09-397-238A-12
13	29	78.4	573	4	US-09-042-709A-19
14	29	78.4	574	1	US-08-140-729A-7
15	29	78.4	574	1	US-08-546-666-7
16	29	78.4	574	2	US-08-916-745-7
17	29	78.4	574	2	US-08-948-569A-4
18	29	78.4	574	2	US-08-663-808-6
19	29	78.4	574	2	US-09-042-929-7
20	29	78.4	574	2	US-08-546-661-7
21	29	78.4	574	2	US-09-042-960-7
22	29	78.4	574	2	US-09-188-469-4
23	29	78.4	574	3	US-09-198-650-7
24	29	78.4	574	3	US-09-332-740-6
25	29	78.4	574	3	US-09-042-913-7
26	29	78.4	574	3	US-09-188-496-6
27	29	78.4	574	3	US-09-042-937-7

28	29	78.4	574	4	US-09-397-238A-4	Sequence 4, Appli
29	29	78.4	574	4	US-09-368-282-6	Sequence 6, Appli
30	29	78.4	574	4	US-09-566-708A-6	Sequence 6, Appli
31	29	78.4	574	4	US-09-042-709A-7	Sequence 7, Appli
32	28	75.7	524	1	US-08-529-654-2	Sequence 2, Appli
33	28	75.7	524	4	US-09-042-709A-20	Sequence 20, Appli
34	28	75.7	524	1	US-08-140-729A-9	Sequence 9, Appli
35	28	75.7	525	1	US-08-546-666-9	Sequence 9, Appli
36	28	75.7	525	2	US-08-916-745-9	Sequence 9, Appli
37	28	75.7	525	2	US-08-948-569A-6	Sequence 6, Appli
38	28	75.7	525	2	US-08-663-808-8	Sequence 8, Appli
39	28	75.7	525	2	US-09-042-929-9	Sequence 9, Appli
40	28	75.7	525	2	US-08-546-661-9	Sequence 9, Appli
41	28	75.7	525	2	US-09-042-960-9	Sequence 9, Appli
42	28	75.7	525	2	US-09-188-469-6	Sequence 6, Appli
43	28	75.7	525	3	US-09-198-650-9	Sequence 9, Appli
44	28	75.7	525	3	US-09-332-740-8	Sequence 8, Appli
45	28	75.7	525	3	US-09-332-740-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-09-461-697-81
; Sequence 81, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Purnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 81
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-81

Query Match 100.0%; Score 37; DB 4; Length 363;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 107 LYQAVATI 114

RESULT 2
US-09-461-697-77
; Sequence 77, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Purnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH

FILE REFERENCE: 10001-005-999
CURRENT APPLICATION NUMBER: US/09/461,697
CURRENT FILING DATE: 1999-12-14
NUMBER OF SEQ ID NOS: 466
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 77
LENGTH: 412
TYPE: PRT
ORGANISM: Homo sapiens
US-09-461-697-77

Query Match 100.0%; Score 37; DB 4; Length 412;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
|||
Db 156 LYQAVATI 163

RESULT 3
US-08-695-301A-30
Sequence 30, Application US/08695301A
Patent No. 6093400

GENERAL INFORMATION:
APPLICANT: DANIEL H. ZIMMERMAN, PREM S. SARIN
TITLE OF INVENTION: Modified HGP-30 Peptides, Conjugates,
TITLE OF INVENTION: Compositions and Methods of Use
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Office of Sherman and Shalloway
STREET: 413 N. Washington Street
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22314

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 1.2 mb storage
MEDIUM TYPE: 3.5 inch, 1.44 mb storage
COMPUTER: Dell System 210; Intel 80 286 Microprocessor
OPERATING SYSTEM: MS DOS 6.22
SOFTWARE: Word Perfect, Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/695,301A
FILING DATE: August 9, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Richard A. Steinberg
REGISTRATION NUMBER: 26,588
REFERENCE/DOCKET NUMBER: CELL-101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 549-2282
TELEFAX: (703) 836-0106
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal fragment
FEATURE:

NAME/KEY:
LOCATION: 75 to 94
OTHER INFORMATION: fragment of p-17 gag protein of
OTHER INFORMATION: HIV-IRF
US-08-695-301A-30

Query Match 81.1%; Score 30; DB 3; Length 20;
Best Local Similarity 75.0%; Pred. No. 1.3;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
|||
Db 3 LYNAVATL 10

RESULT 4

US-08-695-304C-30
Sequence 30, Application US/08695304C
Patent No. 6103239
GENERAL INFORMATION:
APPLICANT: DANIEL H. ZIMMERMAN, PREM S. SARIN
TITLE OF INVENTION: Modified HGP-30 Heteroconjugates,
TITLE OF INVENTION: Compositions and Methods of Use
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Office of Sherman and Shalloway
STREET: 413 N. Washington Street
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22314

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 1.2 mb storage
MEDIUM TYPE: 3.5 inch, 1.44 mb storage
COMPUTER: Dell System 210; Intel 80 286 Microprocessor
OPERATING SYSTEM: MS DOS 6.22
SOFTWARE: Word Perfect, Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/695,304C
FILING DATE: August 9, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Richard A. Steinberg
REGISTRATION NUMBER: 26,588
REFERENCE/DOCKET NUMBER: CELL-102
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 549-2282
TELEFAX: (703) 836-0106
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal fragment
FEATURE:
NAME/KEY:
LOCATION: 75 to 94
OTHER INFORMATION: fragment of p-17 gag protein of
OTHER INFORMATION: HIV-IRF
US-08-695-304C-30

Query Match 81.1%; Score 30; DB 3; Length 20;
Best Local Similarity 75.0%; Pred. No. 1.3;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
|||
Db 3 LYNAVATL 10

RESULT 5

US-09-589-768-30
Sequence 30, Application US/09589768
Patent No. 6268472
GENERAL INFORMATION:

APPLICANT: DANIEL H. ZIMMERMAN, PREM S. SARIN
TITLE OF INVENTION: Modified HGP-30 Peptides, Conjugates,
TITLE OF INVENTION: Compositions and Methods of Use
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Office of Sherman and Shalloway
STREET: 413 N. Washington Street
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22314

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 1.2 mb storage
MEDIUM TYPE: 3.5 inch, 1.44 mb storage
COMPUTER: Dell System 210; Intel 80 286 Microprocessor
OPERATING SYSTEM: MS DOS 6.22
SOFTWARE: Word Perfect, Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/589,768
FILING DATE: Concurrently herewith
ATTORNEY/AGENT INFORMATION:
NAME: Richard A. Steinberg
REGISTRATION NUMBER: 26,588
REFERENCE/DOCKET NUMBER: CELL-101.01
TELEPHONE: (703) 549-2282
TELEFAX: (703) 836-0106
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal fragment
FEATURE:
NAME/KEY:
LOCATION: 75 to 94
OTHER INFORMATION: fragment of p-17 gag protein of HIV-1RF
US-09-589-768-30

Query Match 81.1%; Score 30; DB 4; Length 20;
Best Local Similarity 75.0%; Pred. No. 1.3;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
|||
Db 3 LYNNAVATL 10

RESULT 6
US-09-594-845-30
Sequence 30, Application US/09594845
Patent No. 6287565
GENERAL INFORMATION:
APPLICANT: DANIEL H. ZIMMERMAN, PREM S. SARIN
TITLE OF INVENTION: Modified HGP-30 Heteroconjugates,
TITLE OF INVENTION: Compositions and Methods of Use
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Office of Sherman and Shalloway
STREET: 413 N. Washington Street
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 1.2 mb storage
MEDIUM TYPE: 3.5 inch, 1.44 mb storage
COMPUTER: Dell System 210; Intel 80 286 Microprocessor
OPERATING SYSTEM: MS DOS 6.22
SOFTWARE: Word Perfect, Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/594,845
FILING DATE: Concurrently herewith
ATTORNEY/AGENT INFORMATION:
NAME: Richard A. Steinberg
REGISTRATION NUMBER: 26,588
REFERENCE/DOCKET NUMBER: CELL-102.01
TELEPHONE: (703) 549-2282
TELEFAX: (703) 836-0106
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal fragment
FEATURE:
NAME/KEY:
LOCATION: 75 to 94
OTHER INFORMATION: fragment of p-17 gag protein of HIV-1RF
US-09-594-845-30

Query Match 81.1%; Score 30; DB 4; Length 20;
Best Local Similarity 75.0%; Pred. No. 1.3;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
|||
Db 3 LYNNAVATL 10

RESULT 7
US-08-948-569A-10
Sequence 10, Application US/08948569A
Patent No. 5882926
GENERAL INFORMATION:
APPLICANT: Amara, Susan G
APPLICANT: Arriza, Jeffrey L
APPLICANT: Eliasof, Scott
APPLICANT: Kavanaugh, Michael P
TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes
TITLE OF INVENTION: and Uses
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell-Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,569A
FILING DATE: 10-OCT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5882926nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 93,509-F
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-948-569A-10

Query Match 78.4%; Score 29; DB 2; Length 560;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
|||
Db 383 LYEAAAI 390

RESULT 8

US-09-188-469-10
; Sequence 10, Application US/09188469
; Patent No. 5989825
; GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; APPLICANT: Arriza, Jeffrey L
; APPLICANT: Eliasof, Scott
; APPLICANT: Kavanaugh, Michael P
; TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes
; TITLE OF INVENTION: and Uses
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/188,469
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US/08/948,569
; APPLICATION NUMBER: US/08/948,569
; FILING DATE: 10-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5989825nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,509-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 560 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-188-469-10

Query Match 78.4%; Score 29; DB 2; Length 560;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY

1 LYQAVATI 8

||:||||

Db 383 LYEAVAAI 390

RESULT 9

US-09-397-238A-10
; Sequence 10, Application US/09397238A
; Patent No. 6284505
; GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; APPLICANT: Arriza, Jeffrey L
; APPLICANT: Eliasof, Scott
; APPLICANT: Kavanaugh, Michael P
; TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes
; TITLE OF INVENTION: and Uses
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA

ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/397,238A
; FILING DATE: 16-Sep-1999
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6284505nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,509-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 560 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-397-238A-10
Query Match 78.4%; Score 29; DB 4; Length 560;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LYQAVATI 8
||:||||
Db 383 LYEAVAAI 390
RESULT 10
US-08-948-569A-12
; Sequence 12, Application US/08948569A
; Patent No. 5882926
; GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; APPLICANT: Arriza, Jeffrey L
; APPLICANT: Eliasof, Scott
; APPLICANT: Kavanaugh, Michael P
; TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes
; TITLE OF INVENTION: and Uses
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,569A
; FILING DATE: 10-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5882926nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,509-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids

;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-948-569A-12

Query Match 78.4%; Score 29; DB 2; Length 564;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
||:|||||
DB 385 LYEAVAAI 392

RESULT 11

US-09-188-469-12
; Sequence 12, Application US/09188469
; Patent No. 5989825

;
; GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; APPLICANT: Arriza, Jeffrey L
; APPLICANT: Eliasof, Scott
; APPLICANT: Kavanaugh, Michael P
; TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes
; TITLE OF INVENTION: and Uses
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/188,469
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,569
; FILING DATE: 10-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5989825nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,509-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-188-469-12

Query Match 78.4%; Score 29; DB 2; Length 564;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
||:|||||
DB 385 LYEAVAAI 392

RESULT 12

US-09-397-238A-12
; Sequence 12, Application US/09397238A
; Patent No. 6284505

;
; GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; APPLICANT: Arriza, Jeffrey L
; APPLICANT: Eliasof, Scott
; APPLICANT: Kavanaugh, Michael P
; TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes
; TITLE OF INVENTION: and Uses
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/397,238A
; FILING DATE: 16-Sep-1999
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6284505nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,509-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-397-238A-12

Query Match 78.4%; Score 29; DB 4; Length 564;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
||:|||||
DB 385 LYEAVAAI 392

RESULT 13

US-09-042-709A-19
; Sequence 19, Application US/09042709A
; Patent No. 6458571

;
; GENERAL INFORMATION:
; APPLICANT: State of Oregon
; APPLICANT: Amara, Susan G
; APPLICANT: Arriza, Jeffrey L
; TITLE OF INVENTION: AMINO ACID TRANSPORTERS AND USES
; FILE REFERENCE: 93,509-H (99/145)
; CURRENT APPLICATION NUMBER: US/09/042,709A
; CURRENT FILING DATE: 1998-03-17
; PRIOR APPLICATION NUMBER: 140,729
; PRIOR FILING DATE: 1993-10-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-042-709A-19

Query Match 78.4%; Score 29; DB 4; Length 573;
Best Local Similarity 75.0%; Pred. No. 1e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
||:|||||
Db 402 LYEAVAAI 409

RESULT 14
US-08-140-729A-7
; Sequence 7, Application US/08140729A
; Patent No. 5658782
; GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; APPLICANT: Arriza, Jeffrey L
; TITLE OF INVENTION: Amino Acid Transporters and Uses
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 20 OCT 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5658782nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,509
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-140-729A-7

Query Match 78.4%; Score 29; DB 1; Length 574;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
||:|||||
Db 403 LYEAVAAI 410

RESULT 15
US-08-546-666-7
; Sequence 7, Application US/08546666
; Patent No. 5776774
; GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; APPLICANT: Arriza, Jeffrey L
; TITLE OF INVENTION: Amino Acid Transporters and Uses
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/546,666
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/140,729
; FILING DATE: 20 OCT 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5776774nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,509
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-546-666-7

Query Match 78.4%; Score 29; DB 1; Length 574;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
||:|||||
Db 403 LYEAVAAI 410

Search completed: July 18, 2003, 10:01:23
Job time : 1.60664 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:57:18 ; Search time 1.1564 Seconds
(without alignments)
821.584 Million cell updates/sec

Title: US-09-857-308-5
Perfect score: 37
Sequence: 1 LYQAVATI 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 11875970 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap:*
- 5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap1:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap2:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap3:*
- 13: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pap:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap:*
- 16: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	100.0	363	11	US-09-922-261-81
2	37	100.0	412	11	US-09-922-261-77
3	35	94.6	897	15	US-10-043-487-287
4	32	86.5	273	15	US-10-027-806-56
5	32	86.5	273	15	US-10-034-623-56
6	32	86.5	273	15	US-10-027-801-56
7	31	83.8	269	15	US-10-027-806-24
8	31	83.8	269	15	US-10-034-623-24
9	31	83.8	269	15	US-10-027-801-24
10	29	78.4	481	11	US-09-815-923-8
11	29	78.4	574	12	US-09-748-657-2
12	29	78.4	574	12	US-09-040-736-2
13	29	78.4	574	15	US-10-261-399-7
14	29	78.4	760	12	US-09-759-130B-440
15	29	78.4	760	12	US-09-759-130B-446
16	29	78.4	760	15	US-10-042-431-70

17	29	78.4	760	15	US-10-042-431-76	Sequence 76, Appl
18	28	75.7	223	15	US-10-156-761-9526	Sequence 9526, Ap
19	28	75.7	330	15	US-10-156-761-13865	Sequence 13865, A
20	28	75.7	333	15	US-10-156-761-7860	Sequence 7860, Ap
21	28	75.7	335	12	US-09-940-244-349	Sequence 349, App
22	28	75.7	412	15	US-10-101-464A-905	Sequence 905, App
23	28	75.7	482	15	US-10-101-464A-955	Sequence 955, App
24	28	75.7	525	15	US-10-261-399-9	Sequence 9, Appli
25	28	75.7	712	10	US-09-925-302-678	Sequence 678, App
26	28	75.7	997	15	US-10-101-464A-977	Sequence 977, App
27	27	73.0	137	15	US-10-156-761-12625	Sequence 12625, A
28	27	73.0	183	11	US-09-738-626-6050	Sequence 6050, Ap
29	27	73.0	209	10	US-09-764-870-323	Sequence 323, App
30	27	73.0	209	10	US-09-764-853-546	Sequence 546, App
31	27	73.0	209	15	US-10-125-540-323	Sequence 323, App
32	27	73.0	209	15	US-10-103-313-364	Sequence 364, App
33	27	73.0	339	12	US-09-791-489-2	Sequence 2, Appli
34	27	73.0	380	12	US-09-782-816A-56	Sequence 56, Appl
35	27	73.0	535	10	US-09-815-242-13419	Sequence 13419, A
36	27	73.0	535	11	US-09-844-199-2	Sequence 2, Appli
37	27	73.0	611	15	US-10-074-547-4	Sequence 4, Appli
38	27	73.0	2502	10	US-09-772-316-1	Sequence 1, Appli
39	26	70.3	9	8	US-08-600-483-3	Sequence 3, Appli
40	26	70.3	9	10	US-09-789-720-3	Sequence 3, Appli
41	26	70.3	9	11	US-09-912-787-85	Sequence 85, Appl
42	26	70.3	9	11	US-09-954-392-2	Sequence 2, Appli
43	26	70.3	9	15	US-10-014-326-80	Sequence 80, Appl
44	26	70.3	9	15	US-10-106-487-25	Sequence 25, Appl
45	26	70.3	9	15	US-10-114-823B-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-09-922-261-81
; Sequence 81, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 81
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-81

Query Match 100.0%; Score 37; DB 11; Length 363;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 107 LYQAVATI 114

RESULT 2
US-09-922-261-77

; Sequence 77, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COSENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Purnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 77
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-77

Query Match 100.0%; Score 37; DB 11; Length 412;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 156 LYQAVATI 163

RESULT 3

US-10-043-487-287
; Sequence 287, Application US/10043487
; Publication No. US20030055220A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; APPLICANT: Pierre, LERAIN
; TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptide
; TITLE OF INVENTION: mammalian polypeptides
; FILE REFERENCE: B4778A
; CURRENT APPLICATION NUMBER: US/10/043,487
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/261,130
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 287
; LENGTH: 897
; TYPE: PRT
; ORGANISM: Shigella Flexneri
US-10-043-487-287

Query Match 94.6%; Score 35; DB 15; Length 897;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 770 LYQAVATI 777

RESULT 4

US-10-027-806-56
; Sequence 56, Application US/10027806
; Publication No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.

; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCRP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-56

Query Match 86.5%; Score 32; DB 15; Length 273;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 178 LYQAVATI 185

RESULT 5

US-10-034-623-56
; Sequence 56, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCRP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-034-623-56

Query Match 86.5%; Score 32; DB 15; Length 273;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 178 LYQAVATI 185

RESULT 6

US-10-027-801-56
; Sequence 56, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCRP.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 56
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-801-56

Query Match 86.5%; Score 32; DB 15; Length 273;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 178 LYQAVPTI 185
|||||

RESULT 7
US-10-027-806-24
; Sequence 24, Application US/10027806
; Publication No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-24

Query Match 83.8%; Score 31; DB 15; Length 269;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 174 LYQAVPTV 181
|||||

RESULT 8
US-10-034-623-24
; Sequence 24, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-034-623-24

Query Match 83.8%; Score 31; DB 15; Length 269;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 174 LYQAVPTV 181
|||||

RESULT 9
US-10-027-801-24
; Sequence 24, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-801-24

Query Match 83.8%; Score 31; DB 15; Length 269;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 174 LYQAVPTV 181
|||||

RESULT 10
US-09-815-923-8
; Sequence 8, Application US/09815923
; Publication No. US20020197644A1
; GENERAL INFORMATION:
; APPLICANT: Gill, Sarjeet S.
; APPLICANT: Ross, Linda S.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Use of Insect Cell Membrane Transporters as No. US20020197644A
; FILE REFERENCE: Target Sites for Insecticides
; CURRENT APPLICATION NUMBER: US/09/815,923
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Aedes aegypti
; FEATURE:
; OTHER INFORMATION: glutamate transporter
US-09-815-923-8

Query Match 78.4%; Score 29; DB 11; Length 481;
Best Local Similarity 75.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 353 LYEAVAAI 360
|||||

RESULT 11
US-09-748-657-2
; Sequence 2, Application US/09748657
; Patent No. US20020099025A1

GENERAL INFORMATION:
; APPLICANT: HEYWOOD, JAMES
; TITLE OF INVENTION: TREATMENT OF NEUROLOGICAL DISORDERS
; FILE REFERENCE: 102243-8
; CURRENT APPLICATION NUMBER: US/09/748,657
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,764
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-748-657-2

Query Match 78.4%; Score 29; DB 10; Length 574;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 403 LYEAVAAI 410

RESULT 12
US-09-040-736-2
; Sequence 2, Application US/09040736
; Publication No. US20030060617A1
GENERAL INFORMATION:
; APPLICANT: Rothstein, Jeffery D.
; APPLICANT: Lin, Chieng-Liang Glenn
; APPLICANT: Bristol, Lynn A.
; TITLE OF INVENTION: METHODS FOR DETECTING NEUROLOGICAL
; TITLE OF INVENTION: DISORDERS
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040.736
; FILING DATE: 18-MAR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 1699/48148
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-040-736-2

Query Match 78.4%; Score 29; DB 12; Length 574;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 403 LYEAVAAI 410

RESULT 13
US-10-261-399-7
; Sequence 7, Application US/10261399
; Publication No. US20030125538A1
GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; Arriza, Jeffery L
; TITLE OF INVENTION: Amino Acid Transporters and Uses
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/261,399
; FILING DATE: 01-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/140,729
; FILING DATE: 20 OCT 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. US20030125538A1nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,509
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-261-399-7

Query Match 78.4%; Score 29; DB 15; Length 574;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 403 LYEAVAAI 410

RESULT 14
US-09-759-130B-440
; Sequence 440, Application US/09759130B
; Publication No. US20030022279A1
GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirst, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodearl, Andrew
; APPLICANT: Holtzman, Douglas A

```

; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MPI00-5350NMIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 440
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-759-130B-440

```

```

Query Match          78.4%; Score 29; DB 12; Length 760;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 LYQAVAT 7
        ||| |||
Db      620 LYQCVAT 626

```

```

RESULT 15
US-09-759-130B-446
; Sequence 446, Application US/09/59130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirt, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodearl, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MPI00-5350NMIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159

```

```

; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 446
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-759-130B-446

Query Match          78.4%; Score 29; DB 12; Length 760;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 LYQAVAT 7
        ||| |||
Db      620 LYQCVAT 626

Search completed: July 18, 2003, 10:02:40
Job time : 3.1564 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2003, 07:41:22 ; Search time 399 Seconds
(without alignments)
8846.602 Million cell updates/sec

Title: US-09-857-308-2
Perfect score: 1711
Sequence: 1 acgcgactctgctcagc.....aaaaaaaaaaaaaaaaaaaa 1711

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues
Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq1.*
11: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq3.*
13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1288	75.3	1825	11	US-09-922-261-75
2	1233.2	72.1	1239	11	US-09-922-261-76
3	1092	63.8	1092	11	US-09-922-261-80
4	600	35.1	600	11	US-09-922-261-92
5	510	29.8	510	11	US-09-922-261-96
6	505.8	29.6	573	12	US-09-918-995-14733
7	374.2	21.9	379	12	US-09-918-995-36873
8	337	19.7	491	12	US-09-918-995-31390
9	336	19.6	336	11	US-09-922-261-100
10	293.8	17.2	517	11	US-09-764-847-352
11	293.8	17.2	517	15	US-10-092-154-352
12	230.6	13.5	414	12	US-09-918-995-3654
13	216	12.6	216	11	US-09-922-261-104
14	159	9.3	159	11	US-09-922-261-94
15	141	8.2	141	11	US-09-922-261-112
16	132	7.7	132	11	US-09-922-261-114

17	107	6.3	174	11	US-09-783-590-3427	Sequence 3427, Ap
18	105	6.1	105	11	US-09-922-261-78	Sequence 78, Appl
19	96	5.6	96	11	US-09-922-261-88	Sequence 88, Appl
20	90	5.3	90	11	US-09-922-261-116	Sequence 116, Appl
21	87	5.1	87	11	US-09-922-261-90	Sequence 90, Appl
22	72	4.2	72	11	US-09-922-261-118	Sequence 118, Appl
23	69	4.0	69	11	US-09-922-261-84	Sequence 84, Appl
24	60	3.5	60	11	US-09-922-261-106	Sequence 106, Appl
25	57	3.3	57	11	US-09-922-261-120	Sequence 120, Appl
26	51	3.0	51	11	US-09-922-261-98	Sequence 98, Appl
27	48	2.8	48	11	US-09-922-261-108	Sequence 108, Appl
28	42.6	2.5	1654	10	US-09-925-302-156	Sequence 156, Appl
29	42.6	2.5	712	15	US-10-198-846-8022	Sequence 8022, Ap
30	42.2	2.5	1708	15	US-10-106-698-2124	Sequence 2124, Ap
31	42	2.5	520	11	US-09-834-975-635	Sequence 635, Appl
32	42	2.5	2329	10	US-09-800-729-11	Sequence 11, Appl
33	42	2.5	2355	10	US-09-800-729-27	Sequence 27, Appl
34	42	2.5	2500	15	US-10-102-806-88	Sequence 88, Appl
35	41.8	2.4	903	15	US-10-198-846-4296	Sequence 4296, Ap
36	41.4	2.4	168	11	US-09-964-824A-134	Sequence 134, Appl
37	41.2	2.4	308	11	US-09-960-352-7670	Sequence 7670, Ap
38	41	2.4	1195	15	US-10-180-375-25	Sequence 25, Appl
39	41	2.4	2823	15	US-10-103-313-50	Sequence 50, Appl
40	40.8	2.4	892	15	US-10-198-846-7154	Sequence 7154, Ap
41	40.4	2.4	461	15	US-10-198-846-6011	Sequence 6011, Ap
42	40.2	2.3	549	12	US-09-991-936-1736	Sequence 1736, Ap
43	40	2.3	355	15	US-10-198-846-10174	Sequence 10174, A
44	39.6	2.3	312	12	US-09-907-907A-14	Sequence 14, Appl
45	39.6	2.3	439	12	US-09-918-995-13988	Sequence 13988, A

ALIGNMENTS

RESULT 1
US-09-922-261-75
; Sequence 75, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Thornbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 75
; LENGTH: 1825
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-261-75

Query Match	75.38;	Score	1288;	DB	11;	Length	1825;
Best Local Similarity	99.6%;	Pred. No.	0;				
Mismatches	1291;	Conservative	0;	Mismatches	5;	Indels	0;
Gaps	0;						
Qy	344	AATCTGCAAGACTGGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCC	403				
Db	26	AAATGTTGAGACTGGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCC	85				
Qy	404	TTGATTTGCTCCACGGGAGTTCCGTCGTGGTGAAGTCCATGACCCACCCCTGCACCAA	463				
Db	86	TTGATTTGCTCCACGGGAGTTCCGTCGTGGTGAAGTCCATGACCCACCCCTGCACCAA	145				

Qy 828 CCACAAATCTGGCCCGCGGGCTTTGACTGTGCTAATGAGAGTGTCTGGAGACCCCTAA 887
Db 482 CCACAAATCTGGCCCGCGGGCTTTGACTGTGCTAATGAGAGTGTCTGGAGACCCCTAA 541
Qy 888 CTGATGTGGACATGAGTATGCTTAAAGTTTACCAAGTTGCTGCGTTTGTGCGTGTGAC 947
Db 542 CTGATGTGGACATGAGTATGCTTAAAGTTTACCAAGTTGCTGCGTTTGTGCGTGTGAC 601
Qy 948 GGGAGGCGCGGCTGGGACAGACCTCTTCTGATGTGATGGAGAGGATTTCCATGAAG 1007
Db 602 GGGAGGCGCGGCTGGGACAGACCTCTTCTGATGTGATGGAGAGGATTTCCATGAAG 661
Qy 1008 TGGGTATGTGACGTGTGCTCTCCCTCCAGAAAGTTCTGGCAGCACCACATCAAGACTATC 1067
Db 662 TGGGTATGTGACGTGTGCTCTCCCTCCAGAAAGTTCTGGCAGCACCACATCAAGACTATC 721
Qy 1068 ACAGTTACATGCTACAGATAGTAAGCAACTCTCTGAAGATCAAGGAGGAACTGTGAGGACA 1127
Db 722 ACAGTTACATGCTACAGATAGTAAGCAACTCTCTGAAGATCAAGGAGGAACTGTGAGGACA 781
Qy 1128 CTGAGAAGGCCACAGAGGAGCGCTTAAACCTGTGAAGATCAAGGAGGAACTGTGAGGACA 1187
Db 782 CTGAGAAGGCCACAGAGGAGCGCTTAAACCTGTGAAGATCAAGGAGGAACTGTGAGGACA 841
Qy 1188 TCACTTTTCTGTGATGAGGAGCTGGAGGCTGACCTTGCTTCTGGAGACCACTCACTGC 1247
Db 842 TCACTTTTCTGTGATGAGGAGCTGGAGGCTGACCTTGCTTCTGGAGACCACTCACTGC 901
Qy 1248 CTATGGAGTGCTTGGGGCTCAGAGCAAGCGCTTCCCATCAACCTGGAGGTTGAAGCTT 1307
Db 902 CTATGGAGTGCTTGGGGCTCAGAGCAAGCGCTTCCCATCAACCTGGAGGTTGAAGCTT 961
Qy 1308 CACCACAGGCTTCAAGTGCAGAGGTAATGCTTCTCTCTTTTGGAACTCTGGCCCATGTGA 1367
Db 962 CACCACAGGCTTCAAGTGCAGAGGTAATGCTTCTCTCTTTTGGAACTCTGGCCCATGTGA 1021
Qy 1368 AATGAGGCTCAAGAAAGTGAAGAGGCAATGCTCTGGGCAATGTGCTGGGCAAGT 1427
Db 1022 AATGAGGCTCAAGAAAGTGAAGAGGCAATGCTCTGGGCAATGTGCTGGGCAAGT 1081
Qy 1428 ATCTCTTCAGGAGGCTATGTGAGGATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATG 1487
Db 1082 ATCTCTTCAGGAGGCTATGTGAGGATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATG 1141
Qy 1488 ACTCAGATGACGCTATGTTCCACTTCCACTCAGAGGCTTCACTGGGCTCTCCCTGTTT 1547
Db 1142 ACTCAGATGACGCTATGTTCCACTTCCACTCAGAGGCTTCACTGGGCTCTCCCTGTTT 1201
Qy 1548 TCAACAGGCTCCAGAGAGGATGAGGAAATATAA 1585
Db 1202 TCAACAGGCTCCAGAGAGGATGAGGAAATATAA 1239

RESULT 3

US-09-922-261-80
; Sequence 80, Application US/0992261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14

; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-261-80

Query Match 63.8%; Score 1092; DB 11; Length 1092;
Best Local Similarity 100.0%; Pred. No. 7.1e-312; Indels 0; Gaps 0;
Matches 1091; Conservative 0; Mismatches 0;

Qy 494 ATGCTGGACATCCCTCCAGAGCCATGTAGTCTCACCATCCATACGATTTCAGTTGATTGAG 553
Db 1 ATGCTGGACATCCCTCCAGAGCCATGTAGTCTCACCATCCATACGATTTCAGTTGATTGAG 60
Qy 554 CACAACCGAGCTTTCGCAACCTTATGTCACAGCTCAGCCCGAGAAATCAGCAGCAGACA 613
Db 61 CACAACCGAGCTTTCGCAACCTTATGTCACAGCTCAGCCCGAGAAATCAGCAGCAGACA 120
Qy 614 GAAGGTGTAAAAAAGTGAAGAGAGTGAACCTTCTCCCTCGTGGCTGGGTCACTCTCTCTC 673
Db 121 GAAGGTGTAAAAAAGTGAAGAGAGTGAACCTTCTCCCTCGTGGCTGGGTCACTCTCTCTC 180
Qy 674 CTGTGATGACCTCTCGCTTTTAGATGTGAAGAAATCCCAATGCAACATTCAGATCCGCGAC 733
Db 181 CTGTGATGACCTCTCGCTTTTAGATGTGAAGAAATCCCAATGCAACATTCAGATCCGCGAC 240
Qy 734 AGTGACCCAGAGAGTACCTTTTATCGTGGGAAAGGGGAACTGTGACTGNACTCAGCTGG 793
Db 241 AGTGACCCAGAGAGTACCTTTTATCGTGGGAAAGGGGAACTGTGACTGNACTCAGCTGG 300
Qy 794 CACTCTCTGTGGCAGCTCTCTTACCAGGAGTGGCCACAAATCTGGCCACGCGGCTTT 853
Db 301 CACTCTCTGTGGCAGCTCTCTTACCAGGAGTGGCCACAAATCTGGCCACGCGGCTTT 360
Qy 854 GACTGTCTAATGAGAGTGTCTTGGAGACCCCTAACTGATGTGGCAGATGATGATTCCTT 913
Db 361 GACTGTCTAATGAGAGTGTCTTGGAGACCCCTAACTGATGTGGCAGATGATGATTCCTT 420
Qy 914 AAGTTTACCAAGTTGCTGCGTTTGTGTTGGACCGGGAGCCCGGCTGGGACAGACTCCT 973
Db 421 AAGTTTACCAAGTTGCTGCGTTTGTGTTGGACCGGGAGCCCGGCTGGGACAGACTCCT 480
Qy 974 TTTCTCTGATGTGATGAGCAGAGTATTCATGAAGTGGTATTTGGCAGTGTGCTCTCCCTC 1033
Db 481 TTTCTCTGATGTGATGAGCAGAGTATTCATGAAGTGGTATTTGGCAGTGTGCTCTCCCTC 540
Qy 1034 CAGAAGTTCTGGCAGCACCAGATCAAGGAGTATCAGAGTTACATGTACAGATTAGTAAG 1093
Db 541 CAGAAGTTCTGGCAGCACCAGATCAAGGAGTATCAGAGTTACATGTACAGATTAGTAAG 600
Qy 1094 CAACTCTCTGAAGAATATGAAGGATTTCAATCTCTGAGAGCCACAGAGGAGCTTAA 1153
Db 601 CAACTCTCTGAAGAATATGAAGGATTTCAATCTCTGAGAGCCACAGAGGAGCTTAA 660
Qy 1154 CTTGTGAAGATCAAGGAGGAACTGTGAGGAGACATCACATTTTCTCTGCTGAGGAGCTG 1213
Db 661 CTTGTGAAGATCAAGGAGGAACTGTGAGGAGACATCACATTTTCTCTGCTGAGGAGCTG 720
Qy 1214 GAGGCTGACCTTCTTCTGGAGACCACTGCTGCTTATGGAGTGTGGGGCTCAGAGC 1273
Db 721 GAGGCTGACCTTCTTCTGGAGACCACTGCTGCTTATGGAGTGTGGGGCTCAGAGC 780
Qy 1274 GAACGCTTCCCATCTAACCTGGAGGTTGAAGTTTACCAGAGCTTCAAGTGCAGAGGTA 1333
Db 781 GAACGCTTCCCATCTAACCTGGAGGTTGAAGTTTACCAGAGCTTCAAGTGCAGAGGTA 840
Qy 1334 AATGCTTCTCTCTTTTGGAAATCTGGCCCATGTGAAATGAGCCTCAAGAAAGTGAAGAA 1393
Db 841 AATGCTTCTCTCTTTTGGAAATCTGGCCCATGTGAAATGAGCCTCAAGAAAGTGAAGAA 900
Qy 1394 GGCAATGTCTCTGGGATGCTGCTGGGAGTGTGCTGGGAGTGTGCTTTCAGGAGGCTTATGTGAGGC 1453

Db 901 GCAATGTCCTGGGCAATGGTGTGGGCAATGATGCTTCGAGGAGCCTATGTCAGGC 960
QY 1454 ATGAGTGAAGTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTGCCAC 1513
Db 961 ATGAGTGAAGTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTGCCAC 1020
QY 1514 TCCACTGACAGCCTCATGGGTCCTCCCTGTTTCAACACGCGTGCAGAAAGAGGATG 1573
Db 1021 TCCACTGACAGCCTCATGGGTCCTCCCTGTTTCAACACGCGTGCAGAAAGAGGATG 1080
QY 1574 AGGAAATATAA 1585
Db 1081 AGGAAATATAA 1092

RESULT 4

US-09-922-261-92
; Sequence 92, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-261-92

Query Match 35.1%; Score 600; DB 11; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.2e-166;
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 986 ATGGAGCAGGATTCATGAAGTGGGTATGGCAGTGTGCTCCCTCCAGAAAGTTCTGG 1045
Db 1 ATGGAGCAGGATTCATGAAGTGGGTATGGCAGTGTGCTCCCTCCAGAAAGTTCTGG 60
QY 1046 CAGACCCCATCAGGACTATCAGATTACATGCTACAGATTAGTAAGCAACTCTCTGAA 1105
Db 61 CAGACCCCATCAGGACTATCAGATTACATGCTACAGATTAGTAAGCAACTCTCTGAA 120
QY 1106 GAATATGAAGGATTTCAATCCTCAGAAAGCCACAGAGGCGGTAAACCTGTGAAGATC 1165
Db 121 GAATATGAAGGATTTCAATCCTCAGAAAGCCACAGAGGAGCTAAACCTGTGAAGATC 180
QY 1166 AAGGAGAACCTGTGAGGACATACATTTCTCTGCTAGTAGGAGCTGGAGCGTACCTT 1225
Db 181 AAGGAGAACCTGTGAGGACATACATTTCTCTGCTAGTAGGAGCTGGAGCGTACCTT 240
QY 1226 GCTTCTGGAGACCACTGCTGCTATGGAGTGTGTTGGGCTCAGAGGCAACGCTTCCCA 1285
Db 241 GCTTCTGGAGACCACTGCTGCTATGGAGTGTGTTGGGCTCAGAGGCAACGCTTCCCA 300
QY 1286 TCTAACCTGGAGGTTGAAGCTTCAACACAGGCTTCAAGTCAGAGGTTAAATGCTTCTCT 1345
Db 301 TCTAACCTGGAGGTTGAAGCTTCAACACAGGCTTCAAGTCAGAGGTTAAATGCTTCTCT 360
QY 1346 CTTTGAATCTGCGCCCATGTGAATGGAGCCTCAAGAAAGTGAAGAGGCAATGTCTCT 1405

Db 361 CTTTGAATCTGCGCCCATGTGAATGGAGCCTCAAGAAAGTGAAGAGGCAATGTCTCT 420
QY 1406 GGGCATGTGTGCTGGGCAATGATGCTTCGAGGAGCCTATGTCAGGCAATGAGTGAAGCT 1465
Db 421 GGGCATGTGTGCTGGGCAATGATGCTTCGAGGAGCCTATGTCAGGCAATGAGTGAAGCT 480
QY 1466 GGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTCCCACTCCACTGACAGC 1525
Db 481 GGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTCCCACTCCACTGACAGC 540
QY 1526 CTCATGGGTCCTCCCTGTTTCAACACGCGTGCAGAAAGAGGATGAGGAAATATAA 1585
Db 541 CTCATGGGTCCTCCCTGTTTCAACACGCGTGCAGAAAGAGGATGAGGAAATATAA 600

RESULT 5

US-09-922-261-96
; Sequence 96, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-261-96

Query Match 29.8%; Score 510; DB 11; Length 510;
Best Local Similarity 100.0%; Pred. No. 4.5e-140;
Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1076 ATGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTTGTCATCTCTGAGAAG 1135
Db 1 ATGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTTGTCATCTCTGAGAAG 60
QY 1136 GCCACAGAGGCGCTAAACCTGTGAAGATCAAGGAGAACCTGTGAGCGACATCACTTTT 1195
Db 61 GCCACAGAGGCGCTAAACCTGTGAAGATCAAGGAGAACCTGTGAGCGACATCACTTTT 120
QY 1196 CCTGTCACTGAGGAGCTGGAGCTGACCTTCTTCTGAGACCACTGCTCCTCTATGGGA 1255
Db 121 CCTGTCACTGAGGAGCTGGAGCTGACCTTCTTCTGAGACCACTGCTCCTCTATGGGA 180
QY 1256 GTGCTTGGGCTCAGAGGAGCGCTTCCCATCTAACCTGGAGGTTGAAGCTTCCACAG 1315
Db 181 GTGCTTGGGCTCAGAGGAGCGCTTCCCATCTAACCTGGAGGTTGAAGCTTCCACAG 240
QY 1316 GCTTCAAGTGCAGAGGTAATGCTTCTCTTTGGAATCTGGCCCATGTGAAATGGAG 1375
Db 241 GCTTCAAGTGCAGAGGTAATGCTTCTCTTTGGAATCTGGCCCATGTGAAATGGAG 300
QY 1376 CCTCAAGAAGTGAAGAAGCAATGCTCTTGGGCATGGTGTGCTGGGCGAGTGATGCTTC 1435
Db 301 CCTCAAGAAGTGAAGAAGCAATGCTCTTGGGCATGGTGTGCTGGGCGAGTGATGCTTC 360
QY 1436 GAGGAGCCTATGTCAGGCAATGAGTGAAGCTGGGATTCCTCAGAGCCCTGTGACTCAGAT 1495

Db 361 GAGGAGCCTATGTGAGGATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGAT 420
Qy 1496 AGCAGCTATGTTCCCACTCCACTGACAGCCTCATGGGTCCTCCCTGTTTCAACCCAG 1555
Db 421 AGCAGCTATGTTCCCACTCCACTGACAGCCTCATGGGTCCTCCCTGTTTCAACCCAG 480
Qy 1556 CGTGCAGAGAGAGGATGAGAGAAATATAA 1595
Db 481 CGTGCAGAGAGAGGATGAGAGAAATATAA 510

RESULT 6

US-09-918-995-14733
; Sequence 14733, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14733
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(573)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-14733

Query Match 29.6%; Score 505.8; DB 12; Length 573;
Best Local Similarity 99.6%; Pred. No. 8.4e-139;
Matches 507; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1039 GTTCTGCGACACCGCATCAAGGACTATCAGACTATCAGTATCAGTATGATGAGCAACT 1098
Db 54 GTTCTGCGACACCGCATCAAGGACTATCAGTATCAGTATGATGAGCAACT 113
Qy 1099 CTCTGAAGATATGAAGGATTTGTAATCTGTGAAGGCCACAGAGGACGCTAAACCTGT 1158
Db 114 CTCTGAAGATATGAAGGATTTGTAATCTGTGAAGGCCACAGAGGACGCTAAACCTGT 173
Qy 1159 GAAGATCAAGGAGGAACTGTGAGCGACATCACTTTTCTCTGAGGAGGCTGGAGGC 1218
Db 174 GAAGATCAAGGAGGAACTGTGAGCGACATCACTTTTCTCTGAGGAGGCTGGAGGC 233
Qy 1219 TGACCTTGCTCTGGAGACCACTGAGTCCATGAGGAGTCTTGGGCTCAGAGCGAAG 1278
Db 234 TGACCTTGCTCTGGAGACCACTGAGTCCATGAGGAGTCTTGGGCTCAGAGCGAAG 293
Qy 1279 CTTCCCATCTAACCTGGAGGTTGAAGCTTCACACAGGCTTCAAGTGCAGAGGTAATGC 1338
Db 294 CTTCCCATCTAACCTGGAGGTTGAAGCTTCACACAGGCTTCAAGTGCAGAGGTAATGC 353
Qy 1339 TTCTCTCTTTGGAATCTGGCCCATGTGAAATGAGCCCTCAAGAAAGTGAAGAGGCAA 1398
Db 354 TTCTCTCTTTGGAATCTGGCCCATGTGAAATGAGCCCTCAAGAAAGTGAAGAGGCAA 413
Qy 1399 TGCTCTGGCATGGTGTGCTGGGAGTATGATGCTTCGAGGAGCCCTATGTCAGGATGAG 1458
Db 414 TGCTCTGGCATGGTGTGCTGGGAGTATGATGCTTCGAGGAGCCCTATGTCAGGATGAG 473
Qy 1459 TGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGTATGGTTCCACCTCCAC 1518
Db 474 TGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGTATGGTTCCACCTCCAC 533

Qy 1519 TGACAGCCTCATGGGTCCTCCCTGTTT 1547
Db 534 TGACAGCCTCATGGGTCCTCCCTGTTT 562

RESULT 7

US-09-918-995-36873
; Sequence 36873, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36873
; LENGTH: 379
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-36873

Query Match 21.9%; Score 374.2; DB 12; Length 379;
Best Local Similarity 99.2%; Pred. No. 5.1e-100;
Matches 376; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 738 ACCCAGAGAGTGAATTTATCGTGGGAAAGGGAACTGTGACTGAACCTAGCTGGCACT 797
Db 1 ACCCAGAGAGTGAATTTATCGAGGAAATGTGAACCTGTGACTGAACCTAGCTGGCACT 60
Qy 798 CTTGTGGCAGCTCTCTACCCAGGAGTGGCCCAATCTCGCCCGGCGGCTTTGACT 857
Db 61 CTTGTGGCAGCTCTCTACCCAGGAGTGGCCCAATCTCGCCCGGCGGCTTTGACT 120
Qy 858 GTCTTAATGAGAGTGTCTCGACAGCCTTAATGATGTGGCAGATGATGCTTAAGT 917
Db 121 GTCTTAATGAGAGTGTCTCGACAGCCTTAATGATGTGGCAGATGATGCTTAAGT 180
Qy 918 TTACCAAGTTGCTCGCTTTTGTGTGGACCGGAGGCGCGGCTGGGAGACACTCCTTTTC 977
Db 181 TTACCAAGTTGCTCGCTTTTGTGTGGACCGGAGGCGCGGCTGGGAGACACTCCTTTTC 240
Qy 978 CTGATGTGATGGAGCAGGATTTCCATGAAGTGGGTATTTGGCAGTGTGCTCTCCCTCAGA 1037
Db 241 CTGATGTGATGGAGCAGGATTTCCATGAAGTGGGTATTTGGCAGTGTGCTCTCCCTCAGA 300
Qy 1038 AGTTCTGGCAGCACCCTCAAGGACTATTCACAGTTACATGCTACAGATTAGTAAGCAAC 1097
Db 301 AGTTCTGGCAGCACCCTCAAGGACTATTCACAGTTACATGCTACAGATTAGTAAGCAAC 360
Qy 1098 TCTCTGAAGAATATGAAG 1116
Db 361 TCTCTGAAGAATATGAAG 379

RESULT 8

US-09-918-995-31390
; Sequence 31390, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31390
; LENGTH: 491
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(491)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-31390

Query Match 19.7%; Score 337; DB 12; Length 491;

Best Local Similarity 95.8%; Pred. No. 5.9e-89;

Matches 346; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
QY 1 ACGGATCCTTGCCTCAGGCTCTCGAGGTCCAGACGCCGCCAGCCGCTCTGCGAGC 60
Db 70 ACGGAGCCTTGCCTCAGGCTCTCGAGGTCCAGACTCCGCCAGTCCGCTCTGCGAGC 129
QY 61 CAGCAGTGAATAGTGTGTACTCTCTCTCGGTTCCAGACTCCGCCAGTCCGCTCTTCC 120
Db 130 CACCACTGAATAGTGTGTACTCTCTCTCGGTTCCAGACTCCGCCAGTCCGCTCTTCC 189
QY 121 GCTCCTCTGAACGTCAGGACCTCAGGACCTGTGATGGCGCTCGCGCGGCGGACC 180
Db 190 GCTCCTCTGAACGTCAGGACCTCAGGACCTGTGATGGCGCTCGCGCGGCGGACC 249
QY 181 GTGACCGAGGAACCCCTGGAGGACTTGGGACTTCCCTGGCTCCGCTGTCTTCG 240
Db 250 GTGACCGAGGAACCCCTGGAGGACTTGGGACTTCCCTGGCTCCGCTGTCTTCG 309
QY 241 TGCTCCTTCCGCGCAAGGATCTCATATTATCAGTCTTTTGACCGACACAGAAATGCTGGCA 300
Db 310 CGTCTCTTCCGCGCAAGGATCTCATATTATCAGTCTTTTGACCGACACAGAAATGCTGGCA 369
QY 301 TTTGATAAATGTTTGTGAATCTGAAGACATATGGAACAATGGAACAATCTGCAAGATCTG 360
Db 370 TTTGATAAATGTTTGTGAATCTGAAGACATATGGAACAATGGAACAATCTGCAAGATCTG 429
QY 361 G 361
Db 430 G 430
```

RESULT 9

US-09-922-261-100
; Sequence 100, Application US/0992261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Purnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-261-100

Query Match 19.6%; Score 336; DB 11; Length 336;

```
Best Local Similarity 100.0%; Pred. No. 9.3e-89;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1250 ATGGAGTGTCTGGGCTCAGAGCGAAGCGCTTCCATCTACCTTGGAGTTGAAGCTTCA 1309
Db 1 ATGGAGTGTCTGGGCTCAGAGCGAAGCGCTTCCATCTACCTTGGAGTTGAAGCTTCA 60
QY 1310 CCACAGGCTTCAAGTGCAGAGTAAATGCTTCTCTCTTTTGAATCTGGCCCATGTGAAA 1369
Db 61 CCACAGGCTTCAAGTGCAGAGTAAATGCTTCTCTCTTTTGAATCTGGCCCATGTGAAA 120
QY 1370 ATGGAGCCTCAAGAAAGTGAAGAGGCAATCTCTCTGGGCATGTGTGTCGGCAGTGAT 1429
Db 121 ATGGAGCCTCAAGAAAGTGAAGAGGCAATCTCTCTGGGCATGTGTGTCGGCAGTGAT 180
QY 1430 GTCCTCAGGAGCCTATGTCTAGGCATGAGTGAAGCTGGGATCTCTCAGAGCCCTGATGAC 1489
Db 181 GTCTTCAGGAGCCTATGTCTAGGCATGAGTGAAGCTGGGATCTCTCAGAGCCCTGATGAC 240
QY 1490 TCAGATAGCAGCTATGTTCCACTCCACTCAGAGCCTCATGGGTCTCTCCCTGTTTTC 1549
Db 241 TCAGATAGCAGCTATGTTCCACTCCACTCAGAGCCTCATGGGTCTCTCCCTGTTTTC 300
QY 1550 AACGAGCGCTGCAAGAGAGGATGAGGAAATATAA 1585
Db 301 AACGAGCGCTGCAAGAGAGGATGAGGAAATATAA 336
```

RESULT 10

US-09-764-847-352
; Sequence 352, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 352
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (59)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (482)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (508)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-847-352

Query Match 17.2%;

Best Local Similarity 96.1%; Pred. No. 3.5e-76;

Matches 321; Conservative 1; Mismatches 9; Indels 3; Gaps 2;

```
QY 21 CTCTCGAGTTCACAGACGCCGCCAGCCGCTCTCGACGACGAGTGAATAGTGTGTA 80
Db 17 CCCACGCTCCGACACAGCCGCCAGCCGCTCTCGACGACGAGTGAATAGTGTGTA 76
QY 81 CCTCCTTGTCTCGGTTTCAGGTCCAGACTCCCTCTTTCGGGTGCGCTGAACCTAGGC 140
Db 77 CCTCCTTGTCTCGGTTTCAGGTCCAGACTCCCTCTTTCGGGTGCGCTGAACCTAGGC 136
QY 141 GACCTCAGGACCTGTGATTGGCCCTTGGCCGCGGACCGTGAACGAGGAAACCCCTGG 200
Db 137 GACCTCAGGACCTGTGATTGGCCCTTGGCCGCGGACCGTGAACGAGGAAACCCCTGG 194
```

Query Match	17.2%	Score 293.8	DB 15	Length 517
Best Local Similarity	96.1%	Pred. No. 3.5e-76		
Matches	321	Conservative 1	Mismatches 9	Indels 3
				Gaps 2
QY	21	CTCTCGAGGTCACACAGACGCCGCCAGCCGCTCTGCGACGACGACGATCAATAGTGTGTA	80	
Db	17	CCACGCGTCCGAGACAGCGCGCCAGCCGCTCTGCGACGCANCAAGTGAATAGTGTGTA	76	
QY	81	CTCTCCTGTCTCGGTTCCAGTCCAGACCTCCCCGCTCTCCGGCTGCCCTGAACGTCAGGC	140	
Db	77	CTCTCCTGTCTCGGTTCCAGTCCAGACCTCCCCGCTCTCCGGCTGCCCTGAACGTCAGGC	136	
QY	141	GACCTCAGGACCGTGTGATTGGCGCCTGCGCCGGCGGACCGTGACCGAGGAAACCCCTGG	200	
Db	137	GACCTCAGGACCGTGTGATTGGCGCCTGGCG - GRCGGACCGTGACCGAGGAAACCCCTGG	194	
QY	201	AGGAGCTTGGGCATTCCTTTGGGCTCCGCTGTTCTTCGTGCTCCCTTCGGGCAAGGAT	260	
Db	195	AGGAGCTTGGGCATTCCTTTGGGCTCCGCTGTTCTTCGTGCTCCCTTCGGGCAAGGAT	254	
QY	261	CTCACATTATCAGTCTTTGACCGACACAGAAATGCCTGGCATTTTGATAAATCTTTGTGAA	320	
Db	255	CTCACATTATCAGTCTTTGACCGACACAGAAATGCCTGGCATTTTGATAAATCTTTGTGAA	314	
QY	321	CTTGAAGAGACATAT - GGACAATGAATCTCGAAA	353	
Db	315	CTTGAAGAGACATATGGGACAATGAATCTGGCAA	348	

US-09-918-995-3654

RESULT 13

US-09-922-261-104

; Sequence 104, Application US/09922261

; Patent No. US20020111471A1

; GENERAL INFORMATION:

; APPLICANT: COGENT NEUROSCIENCE, Inc.

; APPLICANT: Lo, Donald C.

; APPLICANT: Barney, Shawn

; APPLICANT: Thomas, Mary Beth

; APPLICANT: Portbury, Stuart D.

; APPLICANT: Puranam, Kasturi

; APPLICANT: Katz, Lawrence C.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING

; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING

; TITLE OF INVENTION: CELL DEATH

; FILE REFERENCE: 10001-005-999

; CURRENT APPLICATION NUMBER: US/09/922,261

; CURRENT FILING DATE: 2001-08-03

; PRIOR APPLICATION NUMBER: US/09/461,697

; PRIOR FILING DATE: 1999-12-14

; NUMBER OF SEQ ID NOS: 466

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 104

; LENGTH: 216

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-922-261-104

Query Match 12.6%; Score 216; DB 11; Length 216;
Best Local Similarity 100.0%; Pred. No. 2.1e-53;
Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1370 ATGGAGCCTCAAGAAACTGAAGAGGCAATGTCCTGGCGATGGTGTGGCGCAGTGAT 1429

```
Db 1 ATGGAGCCTCAGAAAGTGAAGAGCAATGCTCTGGCATGTGTCTGGCAGTGAT 60
QY 1430 GTCTTCGAGGAGCCTATCTCAGGATGAGTGAAGCTGGATTCTCAGAGCCCTGATGAC 1489
Db 61 GTCTTCGAGGAGCCTATCTCAGGATGAGTGAAGCTGGATTCTCAGAGCCCTGATGAC 120
QY 1490 TCAGATAGCAGTATGTTCCACTCCACTGACAGCCCTCATGGGTCTCCCTGTTTC 1549
Db 121 TCAGATAGCAGTATGTTCCACTCCACTGACAGCCCTCATGGGTCTCCCTGTTTC 180
QY 1550 AACAGCCGCTCAAGAGAGGATGAGGAAATATAA 1585
Db 181 AACAGCCGCTCAAGAGAGGATGAGGAAATATAA 216
```

```
RESULT 14
US-09-922-261-94
; Sequence 94, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94
; LENGTH: 159
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-261-94
```

```
Query Match 9.3%; Score 159; DB 11; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.2e-36;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ATGAAGTGGGTATTGGCAGTGTCTCTCCCTCCAGAAAGTTCTGGCAGCAGCCGATCAAGG 1061
Db 1 ATGAAGTGGGTATTGGCAGTGTCTCTCCCTCCAGAAAGTTCTGGCAGCAGCCGATCAAGG 60

QY 1062 ACTATCAGAGTACATGCTACAGATTAGTAAAGCACTCTCTGAAGAATATGAAGGATTG 1121
Db 61 ACTATCAGAGTACATGCTACAGATTAGTAAAGCACTCTCTGAAGAATATGAAGGATTG 120

QY 1122 TCAATCCTCGAAGCCACAGGAGCAGCTAAACCTGTGA 1160
Db 121 TCAATCCTCGAAGCCACAGGAGCAGCTAAACCTGTGA 159
```

```
RESULT 15
US-09-922-261-112
; Sequence 112, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
```

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 112
; LENGTH: 141
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-261-112
```

```
Query Match 8.2%; Score 141; DB 11; Length 141;
Best Local Similarity 100.0%; Pred. No. 2.4e-31;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1445 ATGTCAGGCATGAGTGAAGCTGGGATTCTCAGAGCCCTGATGACTCAGATAGCAGCTAT 1504
Db 1 ATGTCAGGCATGAGTGAAGCTGGGATTCTCAGAGCCCTGATGACTCAGATAGCAGCTAT 60

QY 1505 GGTTCCTCCACTCCACTGACAGCCTCATGGGTCTCTCCCTGTTTTCAACCCAGCGCTGCAAG 1564
Db 61 GGTTCCTCCACTCCACTGACAGCCTCATGGGTCTCTCCCTGTTTTCAACCCAGCGCTGCAAG 120

QY 1565 AAGAGGATGAGGAAATATAA 1585
Db 121 AAGAGGATGAGGAAATATAA 141
```

Search completed: July 18, 2003, 09:55:19
Job time : 403 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	1288	75.3	1825	4	US-09-461-697-75	Sequence 75, Appl
2	1233.2	72.1	1239	4	US-09-461-697-76	Sequence 76, Appl
3	1092	63.8	1092	4	US-09-461-697-80	Sequence 80, Appl
4	600	35.1	600	4	US-09-461-697-92	Sequence 92, Appl
5	510	29.8	510	4	US-09-461-697-96	Sequence 96, Appl
6	336	19.6	336	4	US-09-461-697-100	Sequence 100, Appl
7	216	12.6	216	4	US-09-461-697-104	Sequence 104, Appl
8	159	9.3	159	4	US-09-461-697-94	Sequence 94, Appl
9	141	8.2	141	4	US-09-461-697-112	Sequence 112, Appl
10	132	7.7	132	4	US-09-461-697-114	Sequence 114, Appl
11	105	6.1	105	4	US-09-461-697-78	Sequence 78, Appl
12	96	5.6	96	4	US-09-461-697-88	Sequence 88, Appl
13	90	5.3	90	4	US-09-461-697-116	Sequence 116, Appl
14	87	5.1	87	4	US-09-461-697-90	Sequence 90, Appl
15	72	4.2	72	4	US-09-461-697-118	Sequence 118, Appl
16	69	4.0	69	4	US-09-461-697-84	Sequence 84, Appl
17	60	3.5	60	4	US-09-461-697-106	Sequence 106, Appl
18	57	3.3	57	4	US-09-461-697-120	Sequence 120, Appl
19	51	3.0	51	4	US-09-461-697-98	Sequence 98, Appl
20	48.4	2.8	2502	3	US-09-234-332-1	Sequence 1, Appl
21	48	2.8	48	4	US-09-461-697-108	Sequence 108, Appl
22	43.6	2.5	2089	1	US-08-552-142A-1	Sequence 1, Appl
23	43.6	2.5	2089	4	US-08-910-973-1	Sequence 1, Appl
24	43.6	2.5	2089	4	US-09-499-227-1	Sequence 1, Appl
25	43.6	2.5	2089	5	PCR-US95-05741-1	Sequence 1, Appl
26	39	2.3	45	4	US-09-461-697-124	Sequence 124, Appl
27	38.4	2.2	1831	4	US-09-336-536-15	Sequence 15, Appl

QY 644 CTTCCCTCGTCCCTGGGTACCTCCCTCCTGATGACCTCCTGCTTTAGATTGTAAG 703
Db CTTCCCTCGTCCCTGGGTACCTCCCTCCTGATGACCTCCTGCTTTAGATTGTAAG 385
QY 704 AATCCCAATGACCAATCCAGATCCGACAGTCCGACAGAGAGTACTTTTATCGTGGG 763
Db AATCCCAATGACCAATCCAGATCCGACAGTCCGACAGAGAGTACTTTTATCGTGGG 445
QY 764 AAAGGGAACTGTGACTGAACCTCAGCTGGCAGCTCCTGTGCGCAGCTCCTTACCAGGCA 823
Db AAAGGGAACTGTGACTGAACCTCAGCTGGCAGCTCCTGTGCGCAGCTCCTTACCAGGCA 505
QY 824 GTGGCCACAATCCCTGGCCAGCGGGCTTTGACTGTGCTAATGAGAGTGTCTCGGAGAC 883
Db GTGGCCACAATCCCTGGCCAGCGGGCTTTGACTGTGCTAATGAGAGTGTCTCGGAGAC 565
QY 884 CTAAGTGTGGCAGATGAGTATGCTTAAGTTTACCAAGTGTGCTGCTTTGCTGTG 943
Db CTAAGTGTGGCAGATGAGTATGCTTAAGTTTACCAAGTGTGCTGCTTTGCTGTG 625
QY 944 GACCGGAGGCGCGCTGGGACAGCTCCTTTTCTGTGATGTGAGCAGGATTTCCAT 1003
Db GACCGGAGGCGCGCTGGGACAGCTCCTTTTCTGTGATGTGAGCAGGATTTCCAT 685
QY 1004 GAAGTGGGTATGTCAGTGTCTCCTCCAGAAAGTTCTGGCAGCAGCCGATCAAGGAC 1063
Db GAAGTGGGTATGTCAGTGTCTCCTCCAGAAAGTTCTGGCAGCAGCCGATCAAGGAC 745
QY 1064 TATCAGAGTTACATGCTACAGATAGTAAGCAACTCTGAGAAATATGAAGATGTC 1123
Db TATCAGAGTTACATGCTACAGATAGTAAGCAACTCTGAGAAATATGAAGATGTC 805
QY 1124 AATCCTGAGAGGCGCACAGAGCGCTAAACCTCTGAGAGTCAAGGAGGAACTGTGAGC 1183
Db AATCCTGAGAGGCGCACAGAGCGCTAAACCTCTGAGAGTCAAGGAGGAACTGTGAGC 865
QY 1184 GACATCACTTTTCTGTCAGTGAGAGCTGTGAGGCTGACCTTGTCTGAGAGCACTCA 1243
Db GACATCACTTTTCTGTCAGTGAGAGCTGTGAGGCTGACCTTGTCTGAGAGCACTCA 925
QY 1244 CTGCTATGGAGTGTCTGGGCTCAGAGCAAGCTTCCATCTAACCTGGAGGTTGAA 1303
Db CTGCTATGGAGTGTCTGGGCTCAGAGCAAGCTTCCATCTAACCTGGAGGTTGAA 985
QY 1304 GCTTCAACACAGGCTTCAAGTGCAGAGTAAATGCTTCTCCTTTTGGAACTTGGCCCAT 1363
Db GCTTCAACACAGGCTTCAAGTGCAGAGTAAATGCTTCTCCTTTTGGAACTTGGCCCAT 1045
QY 1364 GTGAAATGGAGCTCAAGAAAGTGAAGAGGCAATGTCTGCGGCAATGTGCTGGG 1423
Db GTGAAATGGAGCTCAAGAAAGTGAAGAGGCAATGTCTGCGGCAATGTGCTGGG 1105
QY 1424 AGTATGTCTTCAGAGGCTTATCTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCT 1483
Db AGTATGTCTTCAGAGGCTTATCTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCT 1165
QY 1484 GATGACTCAGATAGCAGTATGTTCCACTTCCACTGACAGCTCTCATGGGCTCCTCCCT 1543
Db GATGACTCAGATAGCAGTATGTTCCACTTCCACTGACAGCTCTCATGGGCTCCTCCCT 1225
QY 1544 GTTTTCAACAGGCTCAGAGAGAGATGAGGAAATATATAAGGAAAGAGAGGAGATG 1603
Db GTTTTCAACAGGCTCAGAGAGAGATGAGGAAATATATAAGGAAAGAGAGGAGATG 1285
QY 1604 TTTTGTCCAGACCTACTAGACCCCAACAGAAAGTT 1639
Db TTTTGTCCAGACCTACTAGACCCCAACAGAAAGTT 1321

RESULT 2
US-09-461-697-76
; Sequence 76, Application US/09461697
; Patent No. 6277974

GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461.697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-76

Query Match 72.18; Score 1233.2; DB 4; Length 1239;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 348 TGAAGATATCTGGGAGAGATACCAATATCATCAAGCCAGACCAACAGAGTTCTTCG 407
Db 2 TGTGAGATCTGGGAGAGATACCAATATCATCAAGCCAGACCAACAGAGTTCTTCG 61
QY 408 ATTTGCTCCACGGGAGTTCGGTCTGTTGGAAGTCCATGACCCACCCCTGCACCAACCT 467
Db 62 ATTTGCTCCACGGGAGTTCGGTCTGTTGGAAGTCCATGACCCACCCCTGCACCAACCT 121
QY 468 CAGCAACAGCCGAGCCGCCACTATGCTGGACATCCCTCAGAGCCATGATGTCTCA 527
Db 122 CAGCAACAGCCGAGCCGCCACTATGCTGGACATCCCTCAGAGCCATGATGTCTCA 181
QY 528 CCATCCATACGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTG 587
Db 182 CCATCCATACGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTG 241
QY 588 CTAGGCCCCAGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 647
Db 242 CTAGGCCCCAGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 301
QY 648 CCTCGTCCCTGGTCCACCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 707
Db 302 CCTCGTCCCTGGTCCACCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 361
QY 708 CCAATGCACCATTCAGATCCGACAGTCCGACAGTCCGACAGTCCGACAGTCCGACAGT 767
Db 362 CCAATGCACCATTCAGATCCGACAGTCCGACAGTCCGACAGTCCGACAGTCCGACAGT 421
QY 768 GGGAACTGTGACTGAATCAGTGGCAGTCCCTGTCGGCAGTCCCTTACAGGAGTGG 827
Db 422 GGGAACTGTGACTGAATCAGTGGCAGTCCCTGTCGGCAGTCCCTTACAGGAGTGG 481
QY 828 CCAATCTCTGGCCAGCGGGCTTTGACTGTGCTAATGAGAGTGTCTTGAGAGCCTAA 887
Db 482 CCAATCTCTGGCCAGCGGGCTTTGACTGTGCTAATGAGAGTGTCTTGAGAGCCTAA 541
QY 888 CTGATGTGGCAGATGAGTATTGCTTAAGTTTACCAGTTTCTGCGGTTTCTGTGGAC 947
Db 542 CTGATGTGGCAGATGAGTATTGCTTAAGTTTACCAGTTTCTGCGGTTTCTGTGGAC 601
QY 948 GGGAGGCCGCTGGGACAGACTCCTTTCTGATGTGATGGAGAGTATTTCATGAAG 1007
Db 602 GGGAGGCCGCTGGGACAGACTCCTTTCTGATGTGATGGAGAGTATTTCATGAAG 661
QY 1008 TGGGTATTGGCAGTGTCTCCTCCAGAAAGTTCTGGCAGCAGCCGATCAAGGACTATC 1067
Db 662 TGGGTATTGGCAGTGTCTCCTCCAGAAAGTTCTGGCAGCAGCCGATCAAGGACTATC 721

QY 1068 ACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTGTCAATC 1127
Db 722 ACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTGTCAATC 781
QY 1128 CTGAGAAGCCACAGAGGAGCGTAAACCTGTGAAGATCAAGAGGAACCTGTGAGCGACA 1187
Db 782 CTGAGAAGCCACAGAGGAGCGTAAACCTGTGAAGATCAAGAGGAACCTGTGAGCGACA 841
QY 1188 TCACTTTTCTGTCAGTGAAGGAGCTGGAGGCTGACCTTCTCTGGAGACCACTCACTGC 1247
Db 842 TCACTTTTCTGTCAGTGAAGGAGCTGGAGGCTGACCTTCTCTGGAGACCACTCACTGC 901
QY 1248 CTATGGAGTGTCTGGGGCTCAGAGCGAAGCGTTCCTCCATCTAACCTGGAGGTTGAAGCTT 1307
Db 902 CTATGGAGTGTCTGGGGCTCAGAGCGAAGCGTTCCTCCATCTAACCTGGAGGTTGAAGCTT 961
QY 1308 CACCACAGCTTCAAGTGCAGAGGTAAATGCTTCTCTCTTTGGAAATCTGGCCCATGTGA 1367
Db 962 CACCACAGCTTCAAGTGCAGAGGTAAATGCTTCTCTCTTTGGAAATCTGGCCCATGTGA 1021
QY 1368 AAATGAGGCTCAAGAAAGTGAAGAGGCAATGCTCTGGGATGCTGTGGCGAGTG 1427
Db 1022 AAATGAGGCTCAAGAAAGTGAAGAGGCAATGCTCTGGGATGCTGTGGCGAGTG 1081
QY 1428 ATGCTTTCGAGGAGCCTATGTCAAGGCATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATG 1487
Db 1082 ATGCTTTCGAGGAGCCTATGTCAAGGCATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATG 1141
QY 1488 ACTCAGATAGCAGCTATGTTCCCACTCCACTGCAGAGCCTCATGGGGTCTCCCTGTTT 1547
Db 1142 ACTCAGATAGCAGCTATGTTCCCACTCCACTGCAGAGCCTCATGGGGTCTCCCTGTTT 1201
QY 1548 TCAACAGCGCTGCAAGAAGAGGATGAGGAAATATAA 1585
Db 1202 TCAACAGCGCTGCAAGAAGAGGATGAGGAAATATAA 1239

RESULT 3

US-09-461-697-80
; Sequence 80, Application US/09461697
; Patent No. 627974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-80

Query Match 63.8%; Score 1092; DB 4; Length 1092;
Best Local Similarity 100.0%; Pred. No. 7e-302;
Matches 1092; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 494 ATGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCCATACGATTTCAGTTGATTTCAG 553
Db 1 ATGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCCATACGATTTCAGTTGATTTCAG 60
QY 554 CACAACCGAGCTCTTCGCAACCTTATTGCCACAGCTCAGGCCCAAGAAATCAGCAGCAGACA 613

Db 61 CACAACCGAGCTCTTCGCAACCTTATTGCCACAGCTCAGGCCCAAGAAATCAGCAGCAGACA 120
QY 614 GAAGGTGTAAAACACTCAAGAGAGTGAACCTCTTCCTCGTGGCTGGGTCACTCTCCTCTC 673
Db 121 GAAGGTGTAAAACACTCAAGAGAGTGAACCTCTTCCTCGTGGCTGGGTCACTCTCCTCTC 180
QY 674 CCGTATGACCTCCTCGCTTTTAGATTGTAAGAATCCCAATGSCAACATTCCTCAGATCCGGAC 733
Db 181 CCGTATGACCTCCTCGCTTTTAGATTGTAAGAATCCCAATGSCAACATTCCTCAGATCCGGAC 240
QY 734 ATGACCCACAGAGAGTACTTTTATCGTGGAAAGGGGAACCTGTGACTGAACTCAGCTGG 793
Db 241 AGTGACCCACAGAGTACTTTTATCGTGGAAAGGGGAACCTGTGACTGAACTCAGCTGG 300
QY 794 CACTCTCTGCGCAGCTCCTCTACAGAGCAGTGGCCACAATCTGCCCCACCGGGCTTT 853
Db 301 CACTCTCTGCGCAGCTCCTCTACAGAGCAGTGGCCACAATCTGCCCCACCGGGCTTT 360
QY 854 GACTGTGCTTAATGAGAGTGTCTCGGAGACCTTAACCTGATGTGGCACATGAGTATTGCCCTT 913
Db 361 GACTGTGCTTAATGAGAGTGTCTCGGAGACCTTAACCTGATGTGGCACATGAGTATTGCCCTT 420
QY 914 AAGTTTACCAAGTTGCTGCGTTTTCGTGTGGACCGGGAGGCCGGCTGGGACAGACTCTCT 973
Db 421 AAGTTTACCAAGTTGCTGCGTTTTCGTGTGGACCGGGAGGCCGGCTGGGACAGACTCTCT 480
QY 974 TTTCTCTGATGTGATGGAGCAGCTATTCCATGAAGTGGGTATTGGCAGTGTCTCTCCCTC 1033
Db 481 TTTCTCTGATGTGATGGAGCAGCTATTCCATGAAGTGGGTATTGGCAGTGTCTCTCCCTC 540
QY 1034 CAGAAGTGTCTGGCAGCACCGCATCAAGGACTATCACAGTTACATGCTACAGATTAGTAAG 1093
Db 541 CAGAAGTGTCTGGCAGCACCGCATCAAGGACTATCACAGTTACATGCTACAGATTAGTAAG 600
QY 1094 CAACTCTCTGAAGAATATGAAGAAGATTGTCAATCTGTAGAAGGCCACAGAGACCGTAA 1153
Db 601 CAACTCTCTGAAGAATATGAAGAAGATTGTCAATCTGTAGAAGGCCACAGAGACCGTAA 660
QY 1154 CCTGTGAGATCAAGAGGAAACCTGTGAGCGACATCACATTTTCTCTGCTCAGTCAGAGCTG 1213
Db 661 CCTGTGAGATCAAGAGGAAACCTGTGAGCGACATCACATTTTCTCTGCTCAGTCAGAGCTG 720
QY 1214 GAGGTGACCTTCTCTTGGAGACCACTGCTCCCTATGGAGTGTCTTGGGCTCAGAGC 1273
Db 721 GAGGTGACCTTCTCTTGGAGACCACTGCTCCCTATGGAGTGTCTTGGGCTCAGAGC 780
QY 1274 GAACGTTCCCATCTAACCTGGAGGTTGAAGCTTCAACACAGGCTTCAAGTGCAGAGTA 1333
Db 781 GAACGTTCCCATCTAACCTGGAGGTTGAAGCTTCAACACAGGCTTCAAGTGCAGAGTA 840
QY 1334 AATGCTTCTCTCTTTGGAATCTGCCCATGTGAAATGAGGCTTCAAGAAAGTGAAGAA 1393
Db 841 AATGCTTCTCTCTTTGGAATCTGCCCATGTGAAATGAGGCTTCAAGAAAGTGAAGAA 900
QY 1394 GGCAATGTCTCTGGCATGTGTCTGGGACCTGATGTCTTCAGAGAGCCTATGTCAGGC 1453
Db 901 GGCAATGTCTCTGGCATGTGTCTGGGACCTGATGTCTTCAGAGAGCCTATGTCAGGC 960
QY 1454 ATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTCCCTC 1513
Db 961 ATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTCCCTC 1020
QY 1514 TCCACTGCAGCCTCATGGGTCCTCCCTGCTTTTCAACACAGCGCTGCAAGAAAGAGGATG 1573
Db 1021 TCCACTGCAGCCTCATGGGTCCTCCCTGCTTTTCAACACAGCGCTGCAAGAAAGAGGATG 1080
QY 1574 AGGAAATATAA 1585
Db 1081 AGGAAATATAA 1092

```
US-09-461-697-92
; Sequence 92, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461.697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-92

Query Match 35.1%; Score 600; DB 4; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.3e-161;
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 ATGAGCAGGATATCCATGAAGTGGGTATTTGGCAGTGTCTCCCTCCAGAACTTCTGG 1045
DB 1 ATGAGCAGGATATCCATGAAGTGGGTATTTGGCAGTGTCTCCCTCCAGAACTTCTGG 60
QY 1046 CAGCACCACATCAAGGACTATCAGATTACATGTACAGATTAGTACGCACTCTCTGAA 1105
DB 61 CAGCACCACATCAAGGACTATCAGATTACATGTACAGATTAGTACGCACTCTCTGAA 120
QY 1106 GAATATGAAGGATTTGCAATCCCTGAGAAGGCCACAGAGGACGCTAAACCTGTGAAGATC 1165
DB 121 GAATATGAAGGATTTGCAATCCCTGAGAAGGCCACAGAGGACGCTAAACCTGTGAAGATC 180
QY 1166 AAGGAGAACCTGTGAGGACATCACTTTTCTGTGAGTGGAGCTGGAGGCTGACCTT 1225
DB 181 AAGGAGAACCTGTGAGGACATCACTTTTCTGTGAGTGGAGCTGGAGGCTGACCTT 240
QY 1226 GCTTCTGAGACCACTGACTCCCTATGGAGTGTCTGGGCTCAGAGCGACGCTTCCCA 1285
DB 241 GCTTCTGAGACCACTGACTCCCTATGGAGTGTCTGGGCTCAGAGCGACGCTTCCCA 300
QY 1286 TCTAACCTGGAGTTGAAGCTTCCACACAGGCTTCAAGTGAGAGGTTAAATGCTTCTCT 1345
DB 301 TCTAACCTGGAGTTGAAGCTTCCACACAGGCTTCAAGTGAGAGGTTAAATGCTTCTCT 360
QY 1346 CTTTGGAAATCGGCCCATGTGAAAATGGAGCCTCAAGAAAGTGAAGGCAATGTCTCT 1405
DB 361 CTTTGGAAATCGGCCCATGTGAAAATGGAGCCTCAAGAAAGTGAAGGCAATGTCTCT 420
QY 1406 GGGCATGTGTGCTGGGCGAGTGTCTTTCAGGAGGCTATGTTCAGGCAATGAGTGAAGCT 1465
DB 421 GGGCATGTGTGCTGGGCGAGTGTCTTTCAGGAGGCTATGTTCAGGCAATGAGTGAAGCT 480
QY 1466 GGGATTCCTCAGAGCCCTGATGATCAGATAGCAGCTATGTTCCCTACTCCACTGACAGC 1525
DB 481 GGGATTCCTCAGAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 1526 CTCATGGGCTCCCTCCCTCTTTTCAACAGCGCTTCAAGAGAGGATGAGGAAATATAA 1585
DB 541 CTCATGGGCTCCCTCCCTCTTTTCAACAGCGCTTCAAGAGAGGATGAGGAAATATAA 600

RESULT 5
US-09-461-697-96
; Sequence 96, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461.697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-96

Query Match 29.8%; Score 510; DB 4; Length 510;
Best Local Similarity 100.0%; Pred. No. 5.9e-136;
Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1076 ATGCTACAGATTACTAAGCAACTCTCTGAAGAATATGAAGGATTTCAATCCTGAGAAG 1135
DB 1 ATGCTACAGATTACTAAGCAACTCTCTGAAGAATATGAAGGATTTCAATCCTGAGAAG 60
QY 1136 GCCACAGAGACGCTAAACCTGTGAAGATCAAGAGGAACTGTGAGCGACATCAGCTTTT 1195
DB 61 GCCACAGAGACGCTAAACCTGTGAAGATCAAGAGGAACTGTGAGCGACATCAGCTTTT 120
QY 1196 CCTGTAGTGAGGAGCTGGAGGCTGACCTTCTTCTTCTGAGACCACTGCTGCTTATGGGA 1255
DB 121 CCTGTAGTGAGGAGCTGGAGGCTGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 180
QY 1256 GTGCTTGGGCTCAGAGCGAAGCTTCCCATCTAACCTTGAAGGTTGAAGCTTCAACACAG 1315
DB 181 GTGCTTGGGCTCAGAGCGAAGCTTCCCATCTAACCTTGAAGGTTGAAGCTTCAACACAG 240
QY 1316 GCTTCAAGTGCAGAGGTTAAATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1375
DB 241 GCTTCAAGTGCAGAGGTTAAATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 300
QY 1376 CCTCAAGAAAGTGAAGAGGCAATGCTCTGCGGCAATGCTCTGCGGCAATGCTCTGCGG 1435
DB 301 CCTCAAGAAAGTGAAGAGGCAATGCTCTGCGGCAATGCTCTGCGGCAATGCTCTGCGG 360
QY 1436 GAGGAGCCTATGTCAGGCAATGAGTGAAGCTTCTCAGAGCCCTGATGACTCAGAT 1495
DB 361 GAGGAGCCTATGTCAGGCAATGAGTGAAGCTTCTCAGAGCCCTGATGACTCAGAT 420
QY 1496 AGCAGCTATGTTTCCCTACTCCACTGACAGCCTCATGGGGTCTCTCCCTGTTTCAACAC 1555
DB 421 AGCAGCTATGTTTCCCTACTCCACTGACAGCCTCATGGGGTCTCTCCCTGTTTCAACAC 480
QY 1556 CGCTGCAAGAGAGGATGAGGAAATATAA 1585
DB 481 CGCTGCAAGAGAGGATGAGGAAATATAA 510

RESULT 6
US-09-461-697-100
; Sequence 100, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
```

```
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461.697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-100

Query Match      19.6%; Score 336; DB 4; Length 336;
Best Local Similarity 100.0%; Pred. No. 2.2e-86;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1250 ATGGAGTGTGGGGCTCAGAGCAACGCTTCCATCTAACCTGAGAGTTGAAGCTTCA 1309
Db 1 ATGGAGTGTGGGGCTCAGAGCAACGCTTCCATCTAACCTGAGAGTTGAAGCTTCA 60
QY 1310 CCACAGGCTTCAAGTCAGAGGTAATGCTTCTCTCTTTTGGAACTCTGGCCCATGTGAAA 1369
Db 61 CCACAGGCTTCAAGTCAGAGGTAATGCTTCTCTCTTTTGGAACTCTGGCCCATGTGAAA 120
QY 1370 ATGGAGGCTCAAGAAAGTGAAGAGCAATGTCTCTGGGCATGGTGTGGGCAGTGAT 1429
Db 121 ATGGAGGCTCAAGAAAGTGAAGAGCAATGTCTCTGGGCATGGTGTGGGCAGTGAT 180
QY 1430 GTCTTCAGGAGCTATGTTCAGGATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGAC 1489
Db 181 GTCTTCAGGAGCTATGTTCAGGATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGAC 240
QY 1490 TCAGATAGCAGTATGTTCCCACTCCACTGACAGGCTCATGGGGTCTCCCTGTTTC 1549
Db 241 TCAGATAGCAGTATGTTCCCACTCCACTGACAGGCTCATGGGGTCTCCCTGTTTC 300
QY 1550 AACCAGCGCTGCAAGAGAGGATGAGGAAATATAA 1585
Db 301 AACCAGCGCTGCAAGAGAGGATGAGGAAATATAA 336

RESULT 7
US-09-461-697-104
; Sequence 104, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461.697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 104
; LENGTH: 216
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-104

Query Match      12.6%; Score 216; DB 4; Length 216;
Best Local Similarity 100.0%; Pred. No. 4.8e-36;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ATGAAGTGGGTATTGGCAGTGTCTCTCCCTCCAGAGTTCCTGGCAGCACCAGG 1061
Db 1 ATGAAGTGGGTATTGGCAGTGTCTCTCCCTCCAGAGTTCCTGGCAGCACCAGG 60
QY 1062 ACTATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAAATATGAAGGATTG 1121
Db 61 ACTATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAAATATGAAGGATTG 120
QY 1122 TCAATCCTGAGAGGCCACAGAGGACGCTAAACCTGTGA 1160
Db 121 TCAATCCTGAGAGGCCACAGAGGACGCTAAACCTGTGA 159

RESULT 8
US-09-461-697-94
; Sequence 94, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461.697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94
; LENGTH: 159
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-94

Query Match      9.3%; Score 159; DB 4; Length 159;
Best Local Similarity 100.0%; Pred. No. 4.8e-36;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ATGAAGTGGGTATTGGCAGTGTCTCTCCCTCCAGAGTTCCTGGCAGCACCAGG 1061
Db 1 ATGAAGTGGGTATTGGCAGTGTCTCTCCCTCCAGAGTTCCTGGCAGCACCAGG 60
QY 1062 ACTATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAAATATGAAGGATTG 1121
Db 61 ACTATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAAATATGAAGGATTG 120
QY 1122 TCAATCCTGAGAGGCCACAGAGGACGCTAAACCTGTGA 1160
Db 121 TCAATCCTGAGAGGCCACAGAGGACGCTAAACCTGTGA 159

RESULT 9
US-09-461-697-112
; Sequence 112, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
```

```
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 112
; LENGTH: 141
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-112

Query Match
Best Local Similarity 8.2%; Score 141; DB 4; Length 141;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1445 ATGTCAGGCATGAGTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTAT 1504
Db 1 ATGTCAGGCATGAGTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTAT 60

QY 1505 GGTTCCTCCACTGACGTGACAGCCTCATGGGGTCTCCCTGTTTCAACCGCTGCAAG 1564
Db 61 GGTTCCTCCACTGACGTGACAGCCTCATGGGGTCTCCCTGTTTCAACCGCTGCAAG 120

QY 1565 AAGAGGATGAGAAATATAA 1585
Db 121 AAGAGGATGAGAAATATAA 141

RESULT 10
US-09-461-697-114
; Sequence 114, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114
; LENGTH: 132
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-114

Query Match
Best Local Similarity 7.7%; Score 132; DB 4; Length 132;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1454 ATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGCCAC 1513
Db 1 ATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGGTTCCAC 60

QY 1514 TCCACTGACAGCCTCATGGGGTCTCCCTGTTTCAACCGCTGCAAGAAGAGGATG 1573
Db 61 TCCACTGACAGCCTCATGGGGTCTCCCTGTTTCAACCGCTGCAAGAAGAGGATG 120

QY 1574 AGGAAATATAA 1585
Db 1574 AGGAAATATAA 1585
; AAAAAAAAAA
```

```
Db 121 AGGAAATATAA 132

RESULT 11
US-09-461-697-78
; Sequence 78, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 105
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-78

Query Match
Best Local Similarity 6.1%; Score 105; DB 4; Length 105;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 ATGACCCACCCCTGCACCAACCCCTCAGCCACACAGCCGAGCCGCCACTATGCTGGACA 503
Db 1 ATGACCCACCCCTGCACCAACCCCTCAGCCACACAGCCGAGCCGCCACTATGCTGGACA 60

QY 504 TCCCTCAGAGCCATGATGATCTCACCATCCATCAGATTGAGTTGA 548
Db 61 TCCCTCAGAGCCATGATGATCTCACCATCCATCAGATTGAGTTGA 105

RESULT 12
US-09-461-697-88
; Sequence 88, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 88
; LENGTH: 96
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-88

Query Match
Best Local Similarity 5.6%; Score 96; DB 4; Length 96;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 891 ATGTGGCACATGAGTATGCTTAAGTTTACCAAGTTGCTGCTTTTGTGCTGGACCGGG 950
```

Db 1 ATGTGGCACATGAGTATGCTTAAGTTTACCAAGTTGCTGCGCTTTTCTGTGGACCGG 60
QY 951 AGGCCCGCTGGGACAGACTCTTTTCTCTGATGGA 986
Db 61 AGGCCCGCTGGGACAGACTCTTTTCTCTGATGGA 96

RESULT 13
US-09-461-697-116
; Sequence 116, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 116
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-116

Query Match 5.3%; Score 90; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.8e-16;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1485 ATGACTCAGATAGCAGTATGTTCCACTCCACTGACAGCCTCATGGGTCTCCCTG 1544
Db 1 ATGACTCAGATAGCAGTATGTTCCACTCCACTGACAGCCTCATGGGTCTCCCTG 60

QY 1545 TTTTCAACCGCTGCAAGAGGATGA 1574
Db 61 TTTTCAACCGCTGCAAGAGGATGA 90

RESULT 14
US-09-461-697-90
; Sequence 90, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 90
; LENGTH: 87
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-90

Query Match 5.1%; Score 87; DB 4; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.2e-15;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 900 ATGAGTATTCCTTAAGTTTACCAAGTTGCTGCTTTTGTGGACCGGAGCGCGGC 959
Db 1 ATGAGTATTCCTTAAGTTTACCAAGTTGCTGCTTTTGTGGACCGGAGCGCGGC 60

QY 960 TGGGACAGACTCTTTTCTCTGATGGA 986
Db 61 TGGGACAGACTCTTTTCTCTGATGGA 87

RESULT 15
US-09-461-697-118
; Sequence 118, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 118
; LENGTH: 72
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-118

Query Match 4.2%; Score 72; DB 4; Length 72;
Best Local Similarity 100.0%; Pred. No. 2.2e-11;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1503 ATGTTCCCACTCCACTGACAGCCTCATGGGTCTCCCTGTTTCAACGCGCTGCA 1562
Db 1 ATGTTCCCACTCCACTGACAGCCTCATGGGTCTCCCTGTTTCAACGCGCTGCA 60

QY 1563 AGAAGAGGATGA 1574
Db 61 AGAAGAGGATGA 72

Search completed: July 18, 2003, 09:48:24
Job time : 92 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2003, 07:39:02 ; Search time 424 Seconds
(without alignments)
9087.663 Million cell updates/sec

Title: US-09-857-308-2
Perfect score: 1711
Sequence: 1 acgcgatcttgcctcaggc.....aaaaaaaaaaaaaaaaaa 1711

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1711	100.0	1711	21	Human ART-1 nucleo
2	1676.8	98.0	6026	22	Human protein enco
3	1309	76.5	1468	21	Human secreted pro
4	1288	75.3	1825	22	Human cell death p
5	1233.2	72.1	1239	22	Human cell death p
6	1225.6	71.6	2606	24	Human cDNA for nov
7	1092	63.8	1092	22	Human cell death p
8	850.8	49.7	1046	20	Human endometrium
9	600	35.1	600	22	Human cell death p

10	510	29.8	510	22	AAH84181	Human cell death p
11	468	27.4	544	21	AAC08410	Human secreted pro
12	408	23.8	12244	22	AAK89991	Human digestive sy
13	336	19.6	336	22	AAH84183	Human cell death p
14	293.8	17.2	517	23	ABK41954	CDNA encoding nove
15	216	12.6	216	22	AAH84185	Human cell death p
16	159	9.3	159	22	AAH84180	Human cell death p
17	141	8.2	141	22	AAH84189	Human cell death p
18	132	7.7	132	22	AAH84190	Human cell death p
19	109	6.4	2954	22	AAH14356	Human cDNA sequenc
20	108	6.3	554	22	AAH06243	Human cDNA clone (
21	105	6.1	105	22	AAH84172	Human cell death p
22	96	5.6	96	22	AAH84177	Human cell death p
23	90	5.3	90	22	AAH84191	Human cell death p
24	87	5.1	87	22	AAH84178	Human cell death p
25	72	4.2	72	22	AAH84192	Human cell death p
26	69	4.0	69	22	AAH84175	Human cell death p
27	60	3.5	60	22	AAH84186	Human cell death p
28	57	3.3	57	22	AAH84193	Human cell death p
29	51	3.0	51	22	AAH84182	Human cell death p
30	48.4	2.8	1340	18	AAV24018	Human neuro-D gene
31	48.4	2.8	1844	19	AAV24018	Human BHF1 coding
32	48.4	2.8	2502	21	AAK62679	Human NeuroD1 gene
33	48.4	2.8	48	22	AAH84187	Human cell death p
34	46.8	2.7	442	22	AAH83061	Human cell death p
35	44	2.6	1654	21	AAH8137	Human polynucleoti
36	43.6	2.5	2089	16	AAH05513	Lung cancer associ
37	43.6	2.5	2089	18	AAV74887	Murine neurogenic
38	43.6	2.5	2089	19	AAV74887	Mouse neurogenic d
39	42.4	2.5	2572	23	AAH87900	CDNA encoding muri
40	42.2	2.5	168	23	ABV57087	DNA encoding novel
41	42.2	2.5	1708	22	AAH35032	Human prostate exp
42	42	2.5	520	22	AAH60634	Human colon cancer
43	42	2.5	2329	22	AAH97884	Human cancer agent
44	42	2.5	2355	22	AAH97900	Human secreted pro
45	42	2.5	2500	21	AAH21701	Human breast and o

ALIGNMENTS

RESULT 1
AAAG62864
ID AAA62864 standard; DNA; 1711 BP.

AC AAA62864;
XX
DT 26-OCT-2000 (first entry)
XX Human ART-1 nucleotide sequence.

Human; tumour antigen protein; ART-1; HLA antigen; cytostatic; ds;
bound cytotoxic T cell; tumour; cancer; remedy; prevent; diagnose.

Homo sapiens.

WO200032770-A1.

08-JUN-2000.

30-NOV-1999; 99WO-JP06682.

01-DEC-1998; 98JP-0341253.

(SUMU) SUMITOMO PHARM CO LTD.
(ITOH/) ITOH K.

Itoh K, Gomi S;

WPI; 2000-412318/35.
P-PSDB; AAB03880.

Novel tumor antigen protein ART-1, tumor antigen peptide originating

PT from it, their derivatives, and DNAs, applicable in vivo or in vitro as
 XX remedies, preventives and diagnostics for tumors
 PS Claim 2; Page 46-49; 59pp; Japanese.

CC The invention relates to a novel human tumour antigen protein, ART-1.
 CC Included in the invention are polynucleotide sequences encoding the ART-1
 CC protein, and mutated ART-1 proteins which when broken down
 CC intracellularly produce a tumour antigen peptide that can recognise HLA
 CC antigen and bound cytotoxic T cells. Antibodies which specifically
 CC recognise ART-1 and its derivative peptides, are also included in the
 CC invention. ART-1 exhibits cytostatic activity. The tumour antigen
 CC protein, tumour antigen peptide originating from it, their derivatives,
 CC and DNAs are applicable in vivo or in vitro as remedies, preventives and
 CC diagnostics for tumors.
 CC The present sequence DNA encoding the human ART-1 protein.

XX Sequence 1711 BP; 431 A; 452 C; 431 G; 397 T; 0 other;

Query Match	100.0%;	Score 1711;	DB 21;	Length 1711;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1711;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	ACGCGATCCTTGCCCTCAGGCTCTCGAGTCCACAGACCGCCCGCCGCTCTGGCAGC	60		
DB 1	ACGCGATCCTTGCCCTCAGGCTCTCGAGTCCACAGACCGCCCGCCGCTCTGGCAGC	60		
QY 61	CAGCAGTGAATAGTGTGCTACCTCTCTGCTCTCGGTTCCAGTCCAGACCTCCCGCTCTTC	120		
DB 61	CAGCAGTGAATAGTGTGCTACCTCTCTGCTCTCGGTTCCAGTCCAGACCTCCCGCTCTTC	120		
QY 121	GGCTGCCCTGAAGCTCAGGCGACCTCAGGACCTCTGATTTGGGCGCTGCGCGCGGAC	180		
DB 121	GGCTGCCCTGAAGCTCAGGCGACCTCAGGACCTCTGATTTGGGCGCTGCGCGCGGAC	180		
QY 181	GTGACCGAGGAACCCCTGGAGGACTTGGCATCTCTTGGGCTCCGCTGCTCTTCG	240		
DB 181	GTGACCGAGGAACCCCTGGAGGACTTGGCATCTCTTGGGCTCCGCTGCTCTTCG	240		
QY 241	TGCTCTCTTTCGGGGAAGATCTCATTATCAGTCTTTTGACCGACAGAAATGCTGGCA	300		
DB 241	TGCTCTCTTTCGGGGAAGATCTCATTATCAGTCTTTTGACCGACAGAAATGCTGGCA	300		
QY 301	TTTGATTAATGTTTGTGAATCTGAAGACATATGGACATGATCTGCAAGATACTG	360		
DB 301	TTTGATTAATGTTTGTGAATCTGAAGACATATGGACATGATCTGCAAGATACTG	360		
QY 361	GGGAGAGATACCATATCATCAAGCCAGACCAAGAGTTCCTTCGATTGCTCCACG	420		
DB 361	GGGAGAGATACCATATCATCAAGCCAGACCAAGAGTTCCTTCGATTGCTCCACG	420		
QY 421	GGAGTTCCTGCTGGTGAAGTCCATGACCCACCTCTGACCAACCTCAGCACAAGCC	480		
DB 421	GGAGTTCCTGCTGGTGAAGTCCATGACCCACCTCTGACCAACCTCAGCACAAGCC	480		
QY 481	GAAGCCCCCACTATGCTGGACATCCCTTCAGAGCCATGTAGTCTCACCATCCATACGAT	540		
DB 481	GAAGCCCCCACTATGCTGGACATCCCTTCAGAGCCATGTAGTCTCACCATCCATACGAT	540		
QY 541	TCAGTTGATTACAGCACACCGACGCTCTCGCAACCTTATGCCACAGCTCAGGCCGAGAA	600		
DB 541	TCAGTTGATTACAGCACACCGACGCTCTCGCAACCTTATGCCACAGCTCAGGCCGAGAA	600		
QY 601	TCAGCAGCAGACAGAGGTGTAAGAGTGAAGAGTGAAGTTCCTCTGCTGCTCCCTGG	660		
DB 601	TCAGCAGCAGACAGAGGTGTAAGAGTGAAGAGTGAAGTTCCTCTGCTGCTCCCTGG	660		
QY 661	GTCACCTCTCTCCCTGATGACCTCTCTGCTCTTGTAGATTTGAAGATCCCAATGACCAT	720		
DB 661	GTCACCTCTCTCCCTGATGACCTCTCTGCTCTTGTAGATTTGAAGATCCCAATGACCAT	720		
QY 721	CCAGATCCGGCAGACGTACCCAGAGAGTGTATTCGTTGGGGAAGGGGAACTGTGAC	780		

RESULT 2
 AAH99786/C
 ID AAH99786 standard; cdna; 6026 BP.
 XX
 AC AAH99786;
 XX

16-OCT-2001 (first entry).
Human protein encoding cDNA sequence SEQ ID NO:621.
Human; cancer; ulcer; HIV infection; human immunodeficiency virus; antiinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia; antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema; dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection; immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopaenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder; ss.
Homo sapiens.
WO200153455-A2.
26-JUL-2001.
22-DEC-2000; 2000WO-US35017.
23-DEC-1999; 99US-0471275.
21-JAN-2000; 2000US-0488725.
25-APR-2000; 2000US-0552317.
(HYSE-) HYSEQ INC.
Tang YT, Liu C, Drmanac RT;
WPI; 2001-457603/49.
P-PSDB; AAM25845.
isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
Claim 1; Page 647-648; 1217pp; English.
AAH99166 to AAH99904 encode the human proteins given in AAM25225 to AAM25963. The proteins can have activities based on the tissues and cells they are expressed in, such as: antinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis of disorders associated with the activity of a protein e.g. inflammation, rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction, neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal infections, autoimmunity, genetic diseases, haematopoietic disorders, anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers, osteoporosis, severe combined immunodeficiency, eczema, allergic rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression, Alzheimer's disease, Parkinson's disease, neurodegenerative and neurological disorders.
Query Match 98.0%; Score 1676.8; DB 22; Length 6026;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1693; Conservative 0; Mismatches 2; Indels 5; Gaps 1;
1 ACGCGATCCTTGCCCTCAGGCGCTCTCGAGGTCCAGACAGCGCCGCGCTCTCGGACG 60

6023	ACGCGATCCTTGCCCTCAGGCGCTCTCGAGGTCCAGACAGCGCCGCGCTCTCGGACG	5964
61	CACAGTGAATAGTGTGGTACCTCCCTTGTCTCGGTTCCAGTCCAGACCTCCCGTCTTCC	120
5963	CACAGTGAATAGTGTGGTACCTCCCTTGTCTCGGTTCCAGTCCAGACCTCCCGTCTTCC	5904
121	GGTCCCTCAAGCTCAGGCGACCTCAGGACCTGTGATTGGGCGCTCGCGCGGACG	180
5903	GGTCCCTCAAGCTCAGGCGACCTCAGGACCTGTGATTGGGCGCTCGCGCGGACG	5844
181	GTGACGAGAAACCCCTGGAGGACTTGGGCAATCTTGGGCTCGGCTGTGTTCTG	240
5843	GTGACGAGAAACCCCTGGAGGACTTGGGCAATCTTGGGCTCGGCTGTGTTCTG	5784
241	TGCTCCTTTGCG- - - -GGCAAGGATCTCACATTATCATGCTTTCACCGACACAGATGCC	295
5783	TGCTCCTTTGCGTAAAGGCAAGGATCTCACATTATCATGCTTTCACCGACACAGATGCC	5724
296	TGGCATTGTATAATGTTTGTGAACCTTGAAGACATATGGACAATGAATCTGCAAGA	355
5723	TGGCATTGTATAATGTTTGTGAACCTTGAAGACATATGGACAATGAATCTGCAAGA	5664
356	TACTGGGAGAGATACCAATATCATCAAGCCAGACCAACAAGTTCCTTCGATTGCTC	415
5663	TACTGGGAGAGATACCAATATCATCAAGCCAGACCAACAAGTTCCTTCGATTGCTC	5604
416	CCAGGGAGTTCGCTGTGGTGAAGTCCATGACCCACCCCTGACCAACCCCTCAGCAAC	475
5603	CCAGGGAGTTCGCTGTGGTGAAGTCCATGACCCACCCCTGACCAACCCCTCAGCAAC	5544
476	AAGCCGAAGCCGCCACTATGCTGGACATCCCTCAGAGCATGTAGTCTCACCATCAT	535
5543	AAGCCGAAGCCGCCACTATGCTGGACATCCCTCAGAGCATGTAGTCTCACCATCAT	5484
536	ACGATTCAAGTTCAGCACACCGACGCTTCGCGAACCTTATTCGACAGCTCAGGCC	595
5483	ACGATTCAAGTTCAGCACACCGACGCTTCGCGAACCTTATTCGACAGCTCAGGCC	5424
596	CACAATCAGCAGCAGACAGAGGTGTAAAACTGAAAGAGAGTGAACCTCTTCCCTCGTG	655
5423	CAGAATCAGCAGCAGACAGAGGTGTAAAACTGAAAGAGAGTGAACCTCTTCCCTCGTG	5364
656	CCTGGGTACCTCCTCTCCCTGATGACCTCCTGCCTTTAGATTGTAAAGATCCCAATGCA	715
5363	CCTGGGTACCTCCTCTCCCTGATGACCTCCTGCCTTTAGATTGTAAAGATCCCAATGCA	5304
716	CCATTCCAGATCCGGCACAGTACCAGAGAGTGACTTTTATCGTGGGAAAGGGAACTT	775
5303	CCATTCCAGATCCGGCACAGTACCAGAGAGTGACTTTTATCGTGGGAAAGGGAACTT	5244
776	GTGACTGAACCTCAGCTGGGCACTCCTGTGCGCAGCTCCTTACAGGACGTCGCACATC	835
5243	GTGACTGAACCTCAGCTGGGCACTCCTGTGCGCAGCTCCTTACAGGACGTCGCACATC	5184
836	CTGCCCCACCGCGGCTTGTACTGTGCTAATGAGAGTGTCTCTGAGACCTTAATGATGTG	895
5183	CTGCCCCACCGCGGCTTGTACTGTGCTAATGAGAGTGTCTCTGAGACCTTAATGATGTG	5124
896	GCACATGAGTATTCCTTAAAGTTTACCAAGTTCCTGCTGCTTTTCTGTGGACCGGGAGCC	955
5123	GCACATGAGTATTCCTTAAAGTTTACCAAGTTCCTGCTGCTTTTCTGTGGACCGGGAGCC	5064
956	CGGCTGGGACAGACTCCTTTTCTGATGTGATGAGCAGCATGATTCATGAAGTGGGTATT	1015
5063	CGGCTGGGACAGACTCCTTTTCTGATGTGATGAGCAGCATGATTCATGAAGTGGGTATT	5004
1016	GGCAGTGTCTCTCCCTCCAGAAAGTTCCTGGCAGCAGCGCATCAAGGACTATACAGTTAC	1075
5003	GGCAGTGTCTCTCCCTCCAGAAAGTTCCTGGCAGCAGCGCATCAAGGACTATACAGTTAC	4944
1076	ATGCTACAGATTAGTAAGCAACTCTCTGAAAGATATGAAAGGATTGTCAATCTCGAAG	1135

Db 4943 ATGCTACAGTATTAGTAAGCAACTCTCTCAGAANTATGAAAGGATGTCAATCCTGAGAAG 4884
QY 1136 GCCACAGAGGACGCTAAACCTGTGAAGATCAAGGAGGAACCTGTGAGGACATCACATTTT 1195
Db 4883 GCCACAGAGGACGCTAAACCTGTGAAGATCAAGGAGGAACCTGTGAGGACATCACATTTT 4824
QY 1196 CTTGTCAAGTGCAGAGGCTGAGGCTGACCTTCTTCTGAGACCACTGACCTGCTATGGGA 1255
Db 4823 CTTGTCAAGTGCAGAGGCTGAGGCTGACCTTCTTCTGAGACCACTGACCTGCTATGGGA 4764
QY 1256 GTGCTTGGGCTCAGAGGACGCTTCCATCTAACCTGGAGGTTGAAGCTTCAACACAG 1315
Db 4763 GTGCTTGGGCTCAGAGGACGCTTCCATCTAACCTGGAGGTTGAAGCTTCAACACAG 4704
QY 1316 GTTCAAGTGCAGAGGCTGAGGCTTCTCTCTTGGAACTGCGGCCATGTCAAAATGGAG 1375
Db 4703 GTTCAAGTGCAGAGGCTGAGGCTTCTCTCTTGGAACTGCGGCCATGTCAAAATGGAG 4644
QY 1376 CCTCAAGAAAGTGAAGAGGCAATGTCTCTGGGATGTGTCTGGGAGTGTCTGGGAGTGTCTTC 1435
Db 4643 CCTCAAGAAAGTGAAGAGGCAATGTCTCTGGGATGTGTCTGGGAGTGTCTGGGAGTGTCTTC 4584
QY 1436 GAGGAGCTATGTCTCAGGCATGAGTCAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGAT 1495
Db 4583 GAGGAGCTATGTCTCAGGCATGAGTCAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGAT 4524
QY 1496 AGCAGCTATGTCTCCACTCCACTGACAGCCTCATGGGCTCTCCCTGTTTCAACACAG 1555
Db 4523 AGCAGCTATGTCTCCACTCCACTGACAGCCTCATGGGCTCTCCCTGTTTCAACACAG 4464
QY 1556 CGTGCAGAGAGGATGAGGAAATATPAAAGGAAAGAGGGAGATGTTTGTCCAGAC 1615
Db 4463 CGTGCAGAGAGGATGAGGAAATATPAAAGGAAAGAGGGAGATGTTTGTCCAGAC 4404
QY 1616 CTACTAGACCAACAGAAAAGGTTTGTATAGAACTGTTTCTTAAAAATGATTTG 1675
Db 4403 CTACTAGACCAACAGAAAAGGTTTGTATAGAACTGTTTCTTAAAAATGATTTG 4344
QY 1676 ACTCTGTCTTAAAAAAA 1695
Db 4343 ACTCTGTCTTAAACAA 4324
RESULT 3
ID AAC80572
XX AAC80572 standard; cDNA; 1468 BP.
AC AAC80572;
XX
DT 12-FEB-2001 (first entry)
XX Human secreted protein gene 42 SEQ ID NO:52.
DE
DE
KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiac; vasotropic; cerebroprotective;
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; autoimmune disease; hyperproliferative disorder;
KW cardiovascular disorder; cerebrovascular disorder; wound healing;
KW nervous system disorder; aging; chemotaxis; ss.
XX
OS Homo sapiens.
XX
XX WO2000058467-A1.
PN
PD 05-OCT-2000.
XX
PF 22-MAR-2000; 2000WO-US07505.
XX
PR 26-MAR-1999; 99US-0126502.
PR 17-DEC-1999; 99US-0172410.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX

PI Rosen CA, Ruben SM, Komatsoulis G;
XX WPI: 2000-611712/58.
DR P-PSDB; RAB45161.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX
PS Claim 1; Page 373-374; 440pp; English.
XX
XX Polynucleotide sequences AAC80531-C80580 represent cDNA encoding human
CC secreted proteins AAB45120-B45169. Sequences AAB45170-B45225 represent
CC alternative polypeptides encoded by the genes, and amino acid sequences
CC to which they are homologous. The genes and proteins have activities
CC dependent on the tissues and cells in which they are expressed. Examples
CC of their activities include immunosuppressive; antiarthritic;
CC antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;
CC cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
CC fungicide; and ophthalmological. The secreted proteins, polynucleotides,
CC antagonists and agonists may be useful in treating, preventing and/or
CC diagnosing diseases and disorders such as autoimmune diseases
CC e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms
CC of the breast or liver, cardiovascular disorders e.g. cardiac arrest,
CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous
CC system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities. AAC80522-C80530 and AAB45119 represent sequences
CC used in the isolation and characterisation of the genes and proteins of
CC the invention.
XX
SQ Sequence 1468 BP; 399 A; 373 C; 359 G; 337 T; 0 other;
Query Match 76.5%; Score 1309; DB 21; Length 1468;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1328; Conservative 0; Mismatches 10; Indels 3; Gaps 1;
QY 358 CTGGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCCTTCGATTTGCTCCC 417
Db 24 CCGGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCCTTCGATTTGCTCCC 83
QY 418 ACGGAGTTCCTGCTGGTGAAGTCCATGACCCCTCGACCAACCCCTCAGCCACAA 477
Db 84 ACGGAGTTCCTGCTGGTGAAGTCCATGACCCCTCGACCAACCCCTCAGCCACAA 143
QY 478 GCCGAAGCCCCCCTACTATGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCATAC 537
Db 144 GCCGAAGCCCCCCTACTATGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCATAC 203
QY 538 GATTCAGTTGATTCAGACACACCGAGCTTCCTGCAACCTTATTCGCCACAGCTCAGGCCCA 597
Db 204 GATTCAGTTGATTCAGACACACCGAGCTTCCTGCAACCTTATTCGCCACAGCTCAGGCCCA 263
QY 598 GAATCAGCAGCAGACAGAAAGGTGTAATAAAGTGAAGAGAGTGAACCTCTTCCTGCTGCC 657
Db 264 GAATCAGCAGCAGACAGAAAGGTGTAATAAAGTGAAGAGAGTGAACCTCTTCCTGCTGCC 323
QY 658 TGGGTACCTCTCTCCCTGATGACCTCTCTCCCTTTAGATTGTAGAAATCCCAATGCACC 717
Db 324 TGGGTACCTCTCTCCCTGATGACCTCTCTCCCTTTAGATTGTAGAAATCCCAATGCACC 383
QY 718 ATTCAGATCCGGCACAGTACCCAGAGAGTACCTTTTATCGTGGGAAGGGAACCTGT 777
Db 384 ATTCAGATCCGGCACAGTACCCAGAGAGTACCTTTTATCGTGGGAAGGGAACCTGT 443
QY 778 GACTGAACCTCAGCTGGCAGTCTCTGTGCGCAGTCTCTTACAGGAGTGGCCACAACTCT 837
Db 444 GACTGAACCTCAGCTGGCAGTCTCTGTGCGCAGTCTCTTACAGGAGTGGCCACAACTCT 503

QY 838 GGCCACCGGGCTTTGACTGTGCTAATGAGAGTGCCTGGAGACCTTAACATGATGGC 897
|||||
Db 504 GGCCACCGGGCTTTGACTGTGCTAATGAGAGTGCCTGGAGACCTTAACATGATGGC 563
QY 898 ACATGAGTATGCTTAACTTTACCAAGTTGCTGCTTTTGTGCTGGACGGAGCCCG 957
Db 564 ACATGAGTATGCTTAACTTTACCAAGTTGCTGCTTTTGTGCTGGACGGAGCCCG 623
QY 958 GCTGGACAGACTCCCTTTTCC---TGATGTGTGGAGCAGGATTCATCAAGTGGGTAT 1014
|||||
Db 624 GCTGGACAGACTCCCTTTTCCCTTGAATGGAGAGGAGGAGTTCATGAAGTGGGTAT 683
QY 1015 TGCCAGTGTGCTCTCCCTCAGAAAGTTCTGGCAGCACCGCATCAAGGACTATCACAGTTA 1074
|||||
Db 684 TGCCAGTGTGCTCTCCCTCAGAAAGTTCTGGCAGCACCGCATCAAGGACTATCACAGTTA 743
QY 1075 CATGCTACAGATTAGTAAGCACTCTCTGAAGATATGAAGATTGTCAATCCTGAGAA 1134
Db 744 CATGCTACAGATTAGTAAGCACTCTCTGAAGATATGAAGATTGTCAATCCTGAGAA 803
QY 1135 GGCCACAGAGGAGCGCTAAACCTGTGAAGATCAAGGAGGAACTGTGAGCGACATCACTTT 1194
Db 804 GGCCACAGAGGAGCGCTAAACCTGTGAAGATCAAGGAGGAACTGTGAGCGACATCACTTT 863
QY 1195 TCTGTGCTAGTGGAGCTGGAGGCTGACCTTGTCTTGAGAGCACCTGCTGCTATGGG 1254
Db 864 TCTGTGCTAGTGGAGGCTGGAGGCTGACCTTGTCTTGAGAGCACCTGCTGCTATGGG 923
QY 1255 AGTGTGTTGGGCTCAGAGGAGCGCTTCCATCTTAACCTGGAGGTTGAAGCTTCACACA 1314
Db 924 AGTGTGTTGGGCTCAGAGGAGCGCTTCCATCTTAACCTGGAGGTTGAAGCTTCACACA 983
QY 1315 GCTTCAAGTGCAGAGGTAAGTCTCTCTCTTTGGAACTGCGCCATGTGAATGCA 1374
Db 984 GCTTCAAGTGCAGAGGTAAGTCTCTCTCTTTGGAACTGCGCCATGTGAATGCA 1043
QY 1375 GCCTCAAGAAAGTGAAGAGCAATGCTCTCTGGCATGTGTGGGAGTGTGATGCTTT 1434
Db 1044 GCCTCAAGAAAGTGAAGAGCAATGCTCTCTGGCATGTGTGTGGGAGTGTGATGCTTT 1103
QY 1435 CGAGGAGCTATGTTCAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1494
Db 1104 CGAGGAGCTATGTTCAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1163
QY 1495 TAGCAGCTATGTTCCCACTCCACTGACAGCTCATGGGCTCTCCCTGTTTTCACCA 1554
Db 1164 TAGCAGCTATGTTCCCACTCCACTGACAGCTCATGGGCTCTCCCTGTTTTCACCA 1223
QY 1555 GCGCTGCAAGAAGAGGATGAGGAAATATAAAGGAAAGAGGAGATGTTTGTCCAGA 1614
Db 1224 GCGCTGCAAGAAGAGGATGAGGAAATATAAAGGAAAGAGGAGATGTTTGTCCAGA 1283
QY 1615 CCTACTAGACCAACAGAAAGGTTTGTATAGATCTGTTTCTTAAAAATGATTT 1674
Db 1284 CCTACTAGACCAACAGAAAGGTTTGTATAGATCTGTTTCTTAAAAATGATTT 1343
QY 1675 GACTCTGCTTTTAAAAAAA 1695
Db 1344 GACTCTGCTTTTAAACAAA 1364

RESULT 4

AAH84170

XX AAH84170 standard; cDNA; 1825 BP.

AC AAH84170;

XX

DT 21-SEP-2001 (first entry)

XX Human cell death protective cDNA clone CNI-00714, SEQ:75.

DE

XX Cell death protective; apoptosis; necrosis; human; drug screening;

cell death-associated disorder; central nervous system disorder;
psychiatric disorder; neurological disorder; ischaemia-related disorder;
stroke; cerebral infarction; ischaemic encephalopathy;
neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
vascular disease; ophthalmological disorder; diabetic retinopathy;
macular degeneration; hypertension; myocardial infarction;
atherosclerosis; respiratory disorder; asthma; transgenic animal;
chronic obstructive pulmonary disease; neoplastic condition; cancer;
benign tumour; anaemia; gastrointestinal disorder; gastritis;
ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
glomerulonephritis; cystitis; endometriosis; endocrine disorder;
Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS; ss.
Homo sapiens.
WO200145638-A2.
28-JUN-2001.
11-DEC-2000; 2000WO-US33547.
14-DEC-1999; 99US-0461697.
(COGE-) COGENT NEUROSCIENCE INC.
Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
WPI; 2001-390297/41.
P-PSDB; AAG98644, AAG98645, AAG98646, AAG98647, AAG98648, AAG98649,
AAG98650, AAG98651, AAG98652, AAG98653, AAG98654, AAG98655, AAG98656,
AAG98657, AAG98658, AAG98659, AAG98660, AAG98661, AAG98662, AAG98663,
AAG98664, AAG98665, AAG98666, AAG98667, AAG98668, AAG98669, AAG98670,
AAG98671, AAG98672, AAG98673.
Novel protective sequence polynucleotides and polypeptides, used to
identify modulators of their expression and activity, which are used in
to treat central nervous system conditions, diseases and disorders -
Example; Fig 1C; 325pp; English.
Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
protect against cell death (i.e., apoptosis or necrosis). Sequences
AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
while the remaining nucleic acid sequences within the range given above
represent the open reading frames (ORFs) of these cDNA clones. Sequences
AAG98610-AAG98829 represent the polypeptides encoded by the cell death
protective ORFs. The cell death protective cDNA clones are able to
prevent, delay or reverse progression through the apoptotic or necrotic
pathways when injected into a cell predisposed to or undergoing cell
death. The cell death protective nucleic acids and polypeptides can be
used in the diagnosis and treatment of disorders associated with cell
death, and to screen for compounds which modulate their activity or
expression. Such modulators, preferably a small organic molecule, an
antibody, a ribozyme, or an antisense molecule, can also be used to treat
cell death-related diseases. Such diseases include those associated with
the central nervous system including psychiatric or neurological
disorders, especially ischaemia-related conditions such as strokes, and
also includes neurodegenerative disorders such as Alzheimer's disease,
Huntington's disease, or Parkinson's disease. The modulators may also be
used to treat infections such as meningitis, malaria, or trypanosomiasis;
vascular diseases such as ischaemic encephalopathy or cerebral
infarction; eye conditions such as diabetic retinopathy or macular
degeneration; hypertension; myocardial infarction; atherosclerosis;
respiratory conditions such as asthma or chronic obstructive pulmonary
disease; neoplastic conditions such as cancers or benign tumours; blood
cell conditions such as anaemia; gastrointestinal conditions such as
gastritis or ulcerative colitis; liver conditions such as biliary
cirrhosis; kidney disorders such as glomerulonephritis; cystitis;
endometriosis; endocrine disorders such as Grave's disease or Hashimoto's
thyroiditis; skin conditions such as dermatitis or urticaria; or immune

system disorders such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may additionally be used to generate animal models of cell death-associated disorders. The present sequence represents a full-length cell death protective cDNA.

XX
SQ Sequence 1825 BP; 525 A; 447 C; 456 G; 397 T; 0 other;

Query Match 75.3%; Score 1288; DB 22; Length 1825;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 5; Indels 0;

		Matches	1291;	Conservative	0;	Mismatches	5;	Indels	0;	Gaps	0;
Qy	344	AATCTGCAAGATACTCGGGAGAGATACCAATATCATCAAGCCAGACCAACAGAAGTTCC	403								
Db	26	AAATGTTGAGATCTGGGAGAGATACCAATATCATCAAGCCAGACCAACAGAAGTTCC	85								
Qy	404	TTCGATTGCTCCACCGGAGTTCCTGCTGGTGAAGTCCATGACCCACCCCTGCACCA	463								
Db	86	TTCGATTGCTCCACCGGAGTTCCTGCTGGTGAAGTCCATGACCCACCCCTGCACCA	145								
Qy	464	CCCTCAGCCAAACGCGAAGCCGCCACATATGCTGGACATCCCTCAGAGCCATGTAGT	523								
Db	146	CCCTCAGCCAAACGCGAAGCCGCCACATATGCTGGACATCCCTCAGAGCCATGTAGT	205								
Qy	524	CTCACCATCATAGATTCAGTTCAGACACAACCGACGCTCTTCGCAACCTTATTGCC	583								
Db	206	CTCACCATCATAGATTCAGTTCAGTTCAGACACAACCGACGCTCTTCGCAACCTTATTGCC	265								
Qy	584	ACAGCTCAGGCCAGAAATCAGCAGCACAGAGAGTGTAATAACTGAAGAGAGTGAACCT	643								
Db	266	ACAGCTCAGGCCAGAAATCAGCAGCACAGAGAGTGTAATAACTGAAGAGAGTGAACCT	325								
Qy	644	CTTCCCTCGTCCCTGGGTCAACCTCCCTCCCTGATGACCTCCTCGCTTTAGATTGTAAG	703								
Db	326	CTTCCCTCGTCCCTGGGTCAACCTCCCTCCCTGATGACCTCCTCGCTTTAGATTGTAAG	385								
Qy	704	AATCCCNAATGCACCATTCAGATCCGCGACAGTACCCAGAGAGTGACTTTTATCGTGGG	763								
Db	386	AATCCCNAATGCACCATTCAGATCCGCGACAGTACCCAGAGAGTGACTTTTATCGTGGG	445								
Qy	764	AAAGGGGAACCTGTGACTGAACCTGCGACTGCGACTCCTGTGCGGAGCTCCTTACCAAGCA	823								
Db	446	AAAGGGGAACCTGTGACTGAACCTGCGACTGCGACTCCTGTGCGGAGCTCCTTACCAAGCA	505								
Qy	824	GTGGCCACAATCTGGCCACAGGGGGCTTTGACTGTGCTTAATGAGAGTGCTCTGGAGAAC	883								
Db	506	GTGGCCACAATCTGGCCACAGGGGGCTTTGACTGTGCTTAATGAGAGTGCTCTGGAGAAC	565								
Qy	884	CTAACTGATGTGGCACATGAGTATTGCTTTAAAGTTTACCAGTTGCTGCGTTTTCGTGTG	943								
Db	566	CTAACTGATGTGGCACATGAGTATTGCTTTAAAGTTTACCAGTTGCTGCGTTTTCGTGTG	625								
Qy	944	GACCGGAGCCCGGCTGGACAGACTCCTTTTCTGTGATGGAGCAGGTATTCAT	1003								
Db	626	GACCGGAGCCCGGCTGGACAGACTCCTTTTCTGTGATGGAGCAGGTATTCAT	685								
Qy	1004	GAAGTGGGTATTGGCAGTGTCTCTCCTCCAGAAAGTTCTGGCAGCACCCGATCAAGGAC	1063								
Db	686	GAAGTGGGTATTGGCAGTGTCTCTCCTCCAGAAAGTTCTGGCAGCACCCGATCAAGGAC	745								
Qy	1064	TATCACAGTTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTGTC	1123								
Db	746	TATCACAGTTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTGTC	805								
Qy	1124	AATCCTGAGAGGCCACAGAGAGCGCTAAACCTGTGAAGATCAAGGAGGAACCTGTGAGC	1183								
Db	806	AATCCTGAGAGGCCACAGAGAGCGCTAAACCTGTGAAGATCAAGGAGGAACCTGTGAGC	865								
Qy	1184	GACATCACTTTTCTGTCACTGAGGAGCTGGAGGCTGACCTTGCCTCTGAGAGCAGTCA	1243								
Db	866	GACATCACTTTTCTGTCACTGAGGAGCTGGAGGCTGACCTTGCCTCTGAGAGCAGTCA	925								
Qy	1244	CTGCCTATGGAGTGTCTGGGGCTCAGAGGGAACGCTTCCCATCTCAACCTGGAGTTGAA	1303								

Db	926	CTGCCTATGGAGTGTCTGGGGCTCAGAGCAAGCTTCCCATCTACCTGGAGGTTGAA	985
Qy	1304	GCTTCACCACAGGCTTCAAGTGCAGAGGTAATGCTTCTCTCTTTGGAACTCTGGCCCAT	1363
Db	986	GCTTCACCACAGGCTTCAAGTGCAGAGGTAATGCTTCTCTCTTTGGAACTCTGGCCCAT	1045
Qy	1364	GTGAAATGGAGCCTCAAGAAATGAAGAAGCAATGTCTCTGGGATGGTGTCTGTGGC	1423
Db	1046	GTGAAATGGAGCCTCAAGAAATGAAGAAGCAATGTCTCTGGGATGGTGTCTGTGGC	1105
Qy	1424	AGTCATGCTTTCAGGAGGCTATGTCTAGGCATGAGTCAAGCTGGATTCCTCTCAGAGCCCT	1483
Db	1106	AGTGATGCTTTCAGGAGGCTATGTCTAGGCATGAGTCAAGCTGGATTCCTCTCAGAGCCCT	1165
Qy	1484	GATGACTCAGATAGCAGCTATGGTTCCTCCACTCCACTGACACGCTCATGGGGTCTCTCCCT	1543
Db	1166	GATGACTCAGATAGCAGCTATGGTTCCTCCACTCCACTGACACGCTCATGGGGTCTCTCCCT	1225
Qy	1544	GTTTTCAACAGCGCTGCAAGAAGAGATGAGGAAATATATAAGGAAAAAGAGGGAGAT	1603
Db	1226	GTTTTCAACAGCGCTGCAAGAAGAGATGAGGAAATATATAAGGAAAAAGAGGGAGAT	1285
Qy	1604	TTTTGTCCACCTACTAGACCCCAACAGAAAAGTTT	1639
Db	1286	TTTTGTCCACCTACTAGACCCCAACAGAAAAGTTT	1321
RESULT 5			
ID	AAH84171 standard; cDNA; 1239 BP.		
XX	AAH84171;		
XX	21-SEP-2001 (first entry)		
XX	Human cell death protective cDNA clone CNI-00714 ORF1, SEQ.76.		
XX	Cell death protective; apoptosis; necrosis; human; drug screening;		
KW	cell death-associated disorder; central nervous system disorder;		
KW	psychiatric disorder; neurological disorder; ischaemia-related disorder;		
KW	stroke; cerebral infarction; ischaemic encephalopathy;		
KW	neurodegenerative disorder; Alzheimer's disease; Huntington's disease;		
KW	Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;		
KW	vascular disease; ophthalmological disorder; diabetic retinopathy;		
KW	macular degeneration; hypertension; myocardial infarction;		
KW	atherosclerosis; respiratory disorder; asthma; transgenic animal;		
KW	chronic obstructive pulmonary disease; neoplastic condition; cancer;		
KW	benign tumour; anaemia; gastrointestinal disorder; gastritis;		
KW	ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;		
KW	glomerulonephritis; cystitis; endometritis; endocrine disorder;		
KW	Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;		
KW	urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS;		
XX	open reading frame; ORF; ss.		
XX	Homo sapiens.		
OS	WO200145638-A2.		
PN	28-JUN-2001.		
PD	11-DEC-2000; 2000WO-US33547.		
PF	14-DEC-1999; 99US-0461697.		
XX	(COGE-) COGENT NEUROSCIENCE INC.		
PA	Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;		
PI	WPI; 2001-390297/41.		
XX	P-PSDB; AAG98644.		
DR	Novel protective sequence polynucleotides and polypeptides, used to		
PT	identify modulators of their expression and activity, which are used in		

XX AAH84173;
XX
XX
DT 21-SEP-2001 (first entry)
XX
DE Human cell death protective cDNA clone CNI-00714 ORF3, SEQ:80.
XX
XX Cell death protective; apoptosis; necrosis; human; drug screening;
KW cell death-associated disorder; central nervous system disorder;
KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW stroke; cerebral infarction; ischaemic encephalopathy;
KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW vascular degeneration; ophthalmological disorder; diabetic retinopathy;
KW macular degeneration; hypertension; myocardial infarction;
KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS;
XX open reading frame; ORF; ss.
XX
OS Homo sapiens.
XX
XX WO200145638-A2.
XX
XX 28-JUN-2001.
XX
XX 11-DEC-2000; 2000WO-US33547.
XX
XX 14-DEC-1999; 99US-0461697.
XX (COGE-) COGENT NEUROSCIENCE INC.
XX
XX Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
XX
XX WPI; 2001-390297/41.
XX P-PSDB; AAG98646.
XX
XX Novel protective sequence polynucleotides and polypeptides, used to
XX identify modulators of their expression and activity, which are used in
XX to treat central nervous system conditions, diseases and disorders -
XX
XX Claim 2; Fig 6C; 325pp; English.
XX
XX Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
XX protect against cell death (i.e., apoptosis or necrosis). Sequences
XX AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
XX AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
XX while the remaining nucleic acid sequences within the range given above
XX represent the open reading frames (ORFs) of these cDNA clones. Sequences
XX AAG98610-AAG98829 represent the polypeptides encoded by the cell death
XX protective ORFs. The cell death protective cDNA clones are able to
XX prevent, delay or reverse progression through the apoptotic or necrotic
XX pathways when injected into a cell predisposed to or undergoing cell
XX death. The cell death protective nucleic acids and polypeptides can be
XX used in the diagnosis and treatment of disorders associated with cell
XX death, and to screen for compounds which modulate their activity or
XX expression. Such modulators, preferably a small organic molecule, an
XX antibody, a ribozyme, or an antisense molecule, can also be used to treat
XX cell death-related diseases. Such diseases include those associated with
XX the central nervous system including psychiatric or neurological
XX disorders, especially ischaemia-related conditions such as strokes, and
XX also includes neurodegenerative disorders such as Alzheimer's disease,
XX Huntington's disease, or Parkinson's disease. The modulators may also be
XX used to treat infections such as meningitis, malaria, or trypanosomiasis;
XX vascular diseases such as ischaemic encephalopathy or cerebral
XX infarction; eye conditions such as diabetic retinopathy or macular
XX degeneration; hypertension; myocardial infarction; atherosclerosis;
XX respiratory conditions such as asthma or chronic obstructive pulmonary
XX disease; neoplastic conditions such as cancers or benign tumours; blood

CC cell conditions such as anaemia; gastrointestinal conditions such as
CC gastritis or ulcerative colitis; liver conditions such as biliary
CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;
CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's
CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune
CC system disorders such as acquired immunodeficiency syndrome (AIDS). The
CC nucleic acids may additionally be used to generate animal models of
CC cell death-associated disorders. The present sequence represents a
XX cell death protective ORF.
XX
SQ Sequence 1092 BP; 269 A; 280 C; 287 G; 256 T; 0 other;

Query Match 63.8%; Score 1092; DB 22; Length 1092;
Best Local Similarity 100.0%; Pred. No. 5.3e-293;
Matches 1092; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 ATGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATTCATAGGATTGATTCAGTTCAG 553
DB 1 ATGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATTCATAGGATTGATTCAG 60
QY 554 CACAACCGACGCTCTTCGCAACCTTATTGCCACAGCTCAGGCCAGAAATCAGCAGCAGACA 613
DB 61 CACAACCGACGCTCTTCGCAACCTTATTGCCACAGCTCAGGCCAGAAATCAGCAGCAGACA 120
QY 614 GAAGGTGTAAAACTGAAGAGAGTGAACTTCTCCCTCGTGGCTGGGTCACTCCCTCTC 673
DB 121 GAAGGTGTAAAACTGAAGAGAGTGAACTTCTCCCTCGTGGCTGGGTCACTCCCTCTC 180
QY 674 CTTGATGACCTCTCGCTTTTAGATTGTAAAGTCCCAATCCACCATTCAGATCCGGCAC 733
DB 181 CTTGATGACCTCTCGCTTTTAGATTGTAAAGTCCCAATCCACCATTCAGATCCGGCAC 240
QY 734 AGTGACCCAGAGAGTGACTTTTATCGTGGGAAAGGGGAACTGTGACTGAACCTCAGCTGG 793
DB 241 AGTGACCCAGAGAGTGACTTTTATCGTGGGAAAGGGGAACTGTGACTGAACCTCAGCTGG 300
QY 794 CACTCTCTGTCGGCAGCTCTCTTACCAGGAGTGGCCACAATCTGGCCCAACGCGGGCTTT 853
DB 301 CACTCTCTGTCGGCAGCTCTCTTACCAGGAGTGGCCACAATCTGGCCCAACGCGGGCTTT 360
QY 854 GACTGTGCTAATCAGAGTGCTCTGGAGAGCCCTTACTGATGTGGCAGATGATGATTCCTTT 913
DB 361 GACTGTGCTAATCAGAGTGCTCTGGAGAGCCCTTACTGATGTGGCAGATGATGATTCCTTT 420
QY 914 AAGTTTACCAAGTTGCTGCTTTTGTGTCGACCGGAGGCGCGGTGGGACAGACTCCCT 973
DB 421 AAGTTTACCAAGTTGCTGCTTTTGTGTCGACCGGAGGCGCGGTGGGACAGACTCCCT 480
QY 974 TTTCTCTGATGTGGAGCAGGATTTCCATGAAGTGGGTATTTGGCAGTGTGCTCTCCCTC 1033
DB 481 TTTCTCTGATGTGGAGCAGGATTTCCATGAAGTGGGTATTTGGCAGTGTGCTCTCCCTC 540
QY 1034 CAGAAGTTCTGGCAGCAGCCCATCAAGGACTATCAGATTACATGCTACAGATTAGTAAG 1093
DB 541 CAGAAGTTCTGGCAGCAGCCCATCAAGGACTATCAGATTACATGCTACAGATTAGTAAG 600
QY 1094 CAACTCTCTGAAGAATATGAAGGATTGTCAATCTGAGAAGGCCACAGAGGACGCTAAA 1153
DB 601 CAACTCTCTGAAGAATATGAAGGATTGTCAATCTGAGAAGGCCACAGAGGACGCTAAA 660
QY 1154 CTTGTGAAGATCAAGGAGGAACTGTGAGGAGCATCCTTTTCTGTGCTGAGGAGCTG 1213
DB 661 CTTGTGAAGATCAAGGAGGAACTGTGAGGAGCATCCTTTTCTGTGCTGAGGAGCTG 720
QY 1214 GAGGCTGACCTTCTTCTGGAGCAGCTCCTGCTTATGGGAGTGTGGGCTCAGAGC 1273
DB 721 GAGGCTGACCTTCTTCTGGAGCAGCTCCTGCTTATGGGAGTGTGGGCTCAGAGC 780
QY 1274 GAACGCTTCCCATCTAACTGGAGGTTGAAGCTTCCACACAGGCTTCAAGTGCAGAGGTA 1333
DB 781 GAACGCTTCCCATCTAACTGGAGGTTGAAGCTTCCACACAGGCTTCAAGTGCAGAGGTA 840
QY 1334 AATGCTTCTCCTCTTTTGGAAATCTGGCCCATGTGAAAATGGAGCCTCAAGAAAGTGAAGAA 1393

CC vascular diseases such as ischaemic encephalopathy or cerebral
 CC infarction; eye conditions such as diabetic retinopathy or macular
 CC degeneration; hypertension; myocardial infarction; atherosclerosis;
 CC respiratory conditions such as asthma or chronic obstructive pulmonary
 CC disease; neoplastic conditions such as cancers or benign tumours; blood
 CC cell conditions such as anaemia; gastrointestinal conditions such as
 CC gastritis or ulcerative colitis; liver conditions such as biliary
 CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;
 CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's
 CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune
 CC system disorders such as acquired immunodeficiency syndrome (AIDS). The
 CC nucleic acids may additionally be used to generate animal models of
 CC cell death-associated disorders. The present sequence represents a
 CC cell death protective ORF.
 XX
 SQ Sequence 600 BP; 157 A; 138 C; 166 G; 139 T; 0 other;
 Query Match 35.1%; Score 600; DB 22; Length 600;
 Best Local Similarity 100.0%; Pred. No. 2.3e-156;
 Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 986 ATGGAGCAGGTATTCATGAAGTGGTATTTGGCAGTGTGCTCTCCCTCCAGAGTCTGG 1045
 Db 1 ATGGAGCAGGTATTCATGAAGTGGTATTTGGCAGTGTGCTCTCCCTCCAGAGTCTGG 60
 Qy 1046 CAGCACCAGTCAAGGACTATCACAGTTACATGCTACAGATTAGTAAAGCAACTCTCTGAA 1105
 Db 61 CAGCACCAGTCAAGGACTATCACAGTTACATGCTACAGATTAGTAAAGCAACTCTCTGAA 120
 Qy 1106 GAATATGAAGGATTGTCATCTGAGAGGCCACAGAGACGCTAAACCTGTGAAGATC 1165
 Db 121 GAATATGAAGGATTGTCATCTGAGAGGCCACAGAGACGCTAAACCTGTGAAGATC 180
 Qy 1166 AAGCAGGAACCTGTGAGGACATCATTCTCTCAGTGGAGGCTGGAGGCTGACCTT 1225
 Db 181 AAGCAGGAACCTGTGAGGACATCATTCTCTCAGTGGAGGCTGGAGGCTGACCTT 240
 Qy 1226 GCTTCTGGAGACCACTGCTGCTTATGGAGTGTCTGGGGCTCAGAGCGAAGCTTCCCA 1285
 Db 241 GCTTCTGGAGACCACTGCTGCTTATGGAGTGTCTGGGGCTCAGAGCGAAGCTTCCCA 300
 Qy 1286 TCTAACCTGGAGGTTGAAGCTTCAACAGAGCTTCAAGTGCAGAGGTAATGCTTCCTCT 1345
 Db 301 TCTAACCTGGAGGTTGAAGCTTCAACAGAGCTTCAAGTGCAGAGGTAATGCTTCCTCT 360
 Qy 1346 CTTTGGAACTGCGCCCATGTGAATGAGCTCAAGAAATGAGGCTCAAGAAAGTGAAGAGGCAATGCTCT 1405
 Db 361 CTTTGGAACTGCGCCCATGTGAATGAGCTCAAGAAATGAGGCTCAAGAAAGTGAAGAGGCAATGCTCT 420
 Qy 1406 GGGCATGCTGTGCTGGGCACTGATGCTTTCAGAGGAGCTATGTCAGGCATGAGTGAAGCT 1465
 Db 421 GGGCATGCTGTGCTGGGCACTGATGCTTTCAGAGGAGCTATGTCAGGCATGAGTGAAGCT 480
 Qy 1466 GGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGTATGGTTCCCTCCACTCCAGTACAGC 1525
 Db 481 GGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGTATGGTTCCCTCCACTCCAGTACAGC 540
 Qy 1526 CTGATGGGCTCTCCCTGTTTCAACACCGGCTGCAAGAGAGGATGAGGAAATATAA 1585
 Db 541 CTGATGGGCTCTCCCTGTTTCAACACCGGCTGCAAGAGAGGATGAGGAAATATAA 600
 RESULT 10
 AAH84181
 ID AAH84181 standard; cDNA; 510 BP.
 XX
 AC AAH84181;
 XX
 DT 21-SEP-2001 (first entry)
 XX
 DE Human cell death protective cDNA clone CNI-00714 ORF11, SEQ:96.
 XX
 KW Cell death protective; apoptosis; necrosis; human; drug screening;

AAH84179
 ID AAH84179 standard; cDNA; 600 BP.
 XX
 AC AAH84179;
 XX
 DT 21-SEP-2001 (first entry)
 XX
 DE Human cell death protective cDNA clone CNI-00714 ORF9, SEQ:92.
 XX
 KW Cell death protective; apoptosis; necrosis; human; drug screening;
 KW cell death-associated disorder; central nervous system disorder;
 KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
 KW stroke; cerebral infarction; ischaemic encephalopathy;
 KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
 KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
 KW vascular disease; ophthalmological disorder; diabetic retinopathy;
 KW macular degeneration; hypertension; myocardial infarction;
 KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
 KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
 KW benign tumour; anaemia; gastrointestinal disorder; gastritis; disorder;
 KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
 KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
 KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
 KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS;
 KW open reading frame; ORF; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200145638-A2.
 XX
 PD 28-JUN-2001.
 XX
 PE 11-DEC-2000; 2000WO-US33547.
 XX
 PF 14-DEC-1999; 99US-0461697.
 XX
 PR (COGE-) COSENT NEUROSCIENCE INC.
 XX
 PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
 XX
 DR WPI; 2001-390297/41.
 DR P-PSDB; AAG98652.
 XX
 PS Novel protective sequence polynucleotides and polypeptides, used to
 PT identify modulators of their expression and activity, which are used in
 PT to treat central nervous system conditions, diseases and disorders -
 XX
 PS Claim 2; Fig 6I; 325pp; English.
 XX
 CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
 CC protect against cell death (i.e., apoptosis or necrosis). Sequences
 CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
 CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
 CC while the remaining nucleic acid sequences within the range given above
 CC represent the open reading frames (ORFs) of these cDNA clones. Sequences
 CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death
 CC protective ORFs. The cell death protective cDNA clones are able to
 CC prevent, delay or reverse progression through the apoptotic or necrotic
 CC pathways when injected into a cell predisposed to or undergoing cell
 CC death. The cell death protective nucleic acids and polypeptides can be
 CC used in the diagnosis and treatment of disorders associated with cell
 CC death, and to screen for compounds which modulate their activity or
 CC expression. Such modulators, preferably a small organic molecule, an
 CC antibody, a ribozyme, or an antisense molecule, can also be used to treat
 CC cell death-related diseases. Such diseases include those associated with
 CC the central nervous system including psychiatric or neurological
 CC disorders, especially ischaemia-related conditions such as strokes, and
 CC also includes neurodegenerative disorders such as Alzheimer's disease,
 CC Huntington's disease, or Parkinson's disease. The modulators may also be
 CC used to treat infections such as meningitis, malaria, or trypanosomiasis;

PR	14-SEP-2000;	2000US-02330655;
PR	21-SEP-2000;	2000US-02342223;
PR	21-SEP-2000;	2000US-02342274;
PR	25-SEP-2000;	2000US-02349977;
PR	25-SEP-2000;	2000US-02349998;
PR	26-SEP-2000;	2000US-02354884;
PR	27-SEP-2000;	2000US-02358334;
PR	27-SEP-2000;	2000US-02358336;
PR	29-SEP-2000;	2000US-02363277;
PR	29-SEP-2000;	2000US-02363677;
PR	29-SEP-2000;	2000US-02363688;
PR	29-SEP-2000;	2000US-02363690;
PR	29-SEP-2000;	2000US-02363700;
PR	02-OCT-2000;	2000US-02368002;
PR	02-OCT-2000;	2000US-02370377;
PR	02-OCT-2000;	2000US-02370338;
PR	02-OCT-2000;	2000US-02370339;
PR	02-OCT-2000;	2000US-02370404;
PR	13-OCT-2000;	2000US-02399333;
PR	13-OCT-2000;	2000US-02399337;
PR	20-OCT-2000;	2000US-02409600;
PR	20-OCT-2000;	2000US-02412221;
PR	20-OCT-2000;	2000US-02417855;
PR	20-OCT-2000;	2000US-02417866;
PR	20-OCT-2000;	2000US-02417877;
PR	20-OCT-2000;	2000US-02418008;
PR	20-OCT-2000;	2000US-02418009;
PR	20-OCT-2000;	2000US-02418265;
PR	01-NOV-2000;	2000US-02446177;
PR	08-NOV-2000;	2000US-02464774;
PR	08-NOV-2000;	2000US-02464775;
PR	08-NOV-2000;	2000US-02464776;
PR	08-NOV-2000;	2000US-02464777;
PR	08-NOV-2000;	2000US-02464778;
PR	08-NOV-2000;	2000US-02465228;
PR	08-NOV-2000;	2000US-02465332;
PR	08-NOV-2000;	2000US-02465224;
PR	08-NOV-2000;	2000US-02465225;
PR	08-NOV-2000;	2000US-02465226;
PR	08-NOV-2000;	2000US-02465227;
PR	08-NOV-2000;	2000US-02465613;
PR	17-NOV-2000;	2000US-02466110;
PR	17-NOV-2000;	2000US-02466113;
PR	17-NOV-2000;	2000US-02492077;
PR	17-NOV-2000;	2000US-02492080;
PR	17-NOV-2000;	2000US-02492090;
PR	17-NOV-2000;	2000US-02492100;
PR	17-NOV-2000;	2000US-02492111;
PR	17-NOV-2000;	2000US-02492112;
PR	17-NOV-2000;	2000US-02492113;
PR	17-NOV-2000;	2000US-02492114;
PR	17-NOV-2000;	2000US-02492115;
PR	17-NOV-2000;	2000US-02492116;
PR	17-NOV-2000;	2000US-02492117;
PR	17-NOV-2000;	2000US-02492118;
PR	17-NOV-2000;	2000US-02492119;
PR	17-NOV-2000;	2000US-02492120;
PR	17-NOV-2000;	2000US-02492121;
PR	17-NOV-2000;	2000US-02492122;
PR	17-NOV-2000;	2000US-02492123;
PR	17-NOV-2000;	2000US-02492124;
PR	17-NOV-2000;	2000US-02492125;
PR	17-NOV-2000;	2000US-02492126;
PR	17-NOV-2000;	2000US-02492127;
PR	17-NOV-2000;	2000US-02492128;
PR	17-NOV-2000;	2000US-02492129;
PR	17-NOV-2000;	2000US-02492130;
PR	17-NOV-2000;	2000US-02492131;
PR	17-NOV-2000;	2000US-02492132;
PR	17-NOV-2000;	2000US-02492133;
PR	17-NOV-2000;	2000US-02492134;
PR	17-NOV-2000;	2000US-02492135;
PR	17-NOV-2000;	2000US-02492136;
PR	17-NOV-2000;	2000US-02492137;
PR	17-NOV-2000;	2000US-02492138;
PR	17-NOV-2000;	2000US-02492139;
PR	17-NOV-2000;	2000US-02492140;
PR	17-NOV-2000;	2000US-02492141;
PR	17-NOV-2000;	2000US-02492142;
PR	17-NOV-2000;	2000US-02492143;
PR	17-NOV-2000;	2000US-02492144;
PR	17-NOV-2000;	2000US-02492145;
PR	17-NOV-2000;	2000US-02492146;
PR	17-NOV-2000;	2000US-02492147;
PR	17-NOV-2000;	2000US-02492148;
PR	17-NOV-2000;	2000US-02492149;
PR	17-NOV-2000;	2000US-02492150;
PR	17-NOV-2000;	2000US-02492151;
PR	17-NOV-2000;	2000US-02492152;
PR	17-NOV-2000;	2000US-02492153;
PR	17-NOV-2000;	2000US-02492154;
PR	17-NOV-2000;	2000US-02492155;
PR	17-NOV-2000;	2000US-02492156;
PR	17-NOV-2000;	2000US

[illegible]

stroke; cerebral infarction; ischaemic encephalopathy;
neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
vascular disease; ophthalmological disorder; diabetic retinopathy;
macular degeneration; hypertension; myocardial infarction;
atherosclerosis; respiratory disorder; asthma; transgenic animal;
chronic obstructive pulmonary disease; neoplastic condition; cancer;
benign tumour; anaemia; gastrointestinal disorder; gastritis;
ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
glomerulonephritis; cystitis; endometriosis; endocrine disorder;
Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS;
open reading frame; ORF; ss.

Homo sapiens.

W0200145638-A2.

28-JUN-2001.

11-DEC-2000; 2000WO-US33547.

14-DEC-1999; 99US-0461697.

(COGE-) COGENT NEUROSCIENCE INC.

Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;

WPI; 2001-390297/41.

P-PSDB; AAG98656.

Novel protective sequence polynucleotides and polypeptides, used to
identify modulators of their expression and activity, which are used in
to treat central nervous system conditions, diseases and disorders -

Claim 2; Fig 6M; 325pp; English.

Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
protect against cell death (i.e., apoptosis or necrosis). Sequences
AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
while the remaining nucleic acid sequences within the range given above
represent the open reading frames (ORFs) of these cDNA clones. Sequences
AAG98610-AAG98829 represent the polypeptides encoded by the cell death
protective ORFs. The cell death protective cDNA clones are able to
prevent, delay or reverse progression through the apoptotic or necrotic
pathways when injected into a cell predisposed to or undergoing cell
death. The cell death protective nucleic acids and polypeptides can be
used in the diagnosis and treatment of disorders associated with cell
death, and to screen for compounds which modulate their activity or
expression. Such modulators, preferably a small organic molecule, an
antibody, a ribozyme, or an antisense molecule, can also be used to treat
cell death-related diseases. Such diseases include those associated with
the central nervous system including psychiatric or neurological
disorders, especially ischaemia-related conditions such as strokes, and
also includes neurodegenerative disorders such as Alzheimer's disease,
Huntington's disease, or Parkinson's disease. The modulators may also be
used to treat infections such as meningitis, malaria, or trypanosomiasis;
vascular diseases such as ischaemic encephalopathy or cerebral
infarction; eye conditions such as diabetic retinopathy or macular
degeneration; hypertension; myocardial infarction; atherosclerosis;
respiratory conditions such as asthma or chronic obstructive pulmonary
disease; neoplastic conditions such as cancers or benign tumours; blood
cell conditions such as anaemia; gastrointestinal conditions such as
gastritis or ulcerative colitis; liver conditions such as biliary
cirrhosis; kidney disorders such as glomerulonephritis; cystitis;
endometriosis; endocrine disorders such as Grave's disease or Hashimoto's
thyroiditis; skin conditions such as dermatitis or urticaria; or immune
system disorders such as acquired immunodeficiency syndrome (AIDS). The
nucleic acids may additionally be used to generate animal models of
cell death-associated disorders. The present sequence represents a
cell death protective ORF.

SQ Sequence 336 BP; 84 A; 78 C; 96 G; 78 T; 0 other;
Query Match 19.6%; Score 336; DB 22; Length 336;
Best Local Similarity 100.0%; Pred. No. 4.2e-83;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1250 ATGGGAGTGTGGGGCTCAGAGCGAACGCTTCCCATCTAACCTGGAGGTTGAAGCTTCA 1309
DB 1 ATGGGAGTGTGGGGCTCAGAGCGAACGCTTCCCATCTAACCTGGAGGTTGAAGCTTCA 60
QY 1310 CCACAGGCTCAAGTCAGAGGTAATGCTTCTCTCTTTGGGAATCTGCCCATGTGAAA 1369
DB 61 CCACAGGCTCAAGTCAGAGGTAATGCTTCTCTCTTTGGGAATCTGCCCATGTGAAA 120
QY 1370 ATGGAGCCTCAAGAAAGTGAAGAGCAATGCTCTGGGCATGCTGTGGGCAGTGTAT 1429
DB 121 ATGGAGCCTCAAGAAAGTGAAGAGCAATGCTCTGGGCATGCTGTGGGCAGTGTAT 180
QY 1430 GTCTTCGAGGAGCCTATGTGAGGCATGTAGTGAAGCTGGGATTCCTCAGAGCCCTGATGAC 1489
DB 181 GTCTTCGAGGAGCCTATGTGAGGCATGTAGTGAAGCTGGGATTCCTCAGAGCCCTGATGAC 240
QY 1490 TCAGATAGCAGCTATGGTCCCACTCCCACTGACAGCCTCATGGGGTCTTCCCTGTTTC 1549
DB 241 TCAGATAGCAGCTATGGTCCCACTCCCACTGACAGCCTCATGGGGTCTTCCCTGTTTC 300
QY 1550 AACCCAGCGCTGCAAGAGAGGAGGATGAGGAAAAATATAA 1585
DB 301 AACCCAGCGCTGCAAGAGAGGAGGATGAGGAAAAATATAA 336
RESULT 14
ABK41954
ID ABK41954 standard; CDNA; 517 BP.
XX AC ABK41954;
XX 21-MAY-2002 (first entry)
XX CDNA encoding novel human connective tissue related polypeptide #342.
XX Human; connective tissue related disorder; cancer; gene therapy;
XX cytotstatic; gene; ss.
XX Homo sapiens.
XX WO200155343-A1.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01322.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-565190/63.
DR P-PSDB; AAU86776.
XX
PT Nucleic acid encoding novel connective tissue associated polypeptides,
PT used in diagnosing, preventing, treating or ameliorating a disorder
PT such as cancer or rheumatoid arthritis -
XX
PS Claim 4; SEQ ID No 352; 673pp; English.
XX
CC The present invention relates to the isolation of novel human connective
CC tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide
CC (cDNA and genomic) sequences encoding them. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of diseases associated with connective tissue(s), including
CC cancer. The polynucleotide sequences of the invention are also useful
CC in gene therapy. ABK41613-ABK42101 represent cDNA sequences encoding
CC the novel human connective tissue related polypeptides.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 517 BP; 108 A; 140 C; 132 G; 133 T; 4 other;
Query Match 17.2%; Score 293.8; DB 23; Length 517;
Best Local Similarity 96.1%; Pred. No. 2.8e-71;
Matches 321; Conservative 1; Mismatches 9; Indels 3; Gaps 2;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:57:18 ; Search time 59.8436 Seconds
(without alignments)
821.584 Million cell updates/sec

Title: US-09-857-308-1

Perfect score: 2175

Sequence: 1 MNLQRYWGEIPISSQTNRS.....SLMGSPVFNQRCKMRKI 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/FCI_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep1.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep2.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep3.*
- 13: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2155	99.1	412	11	US-09-922-261-77
2	1898	87.3	363	11	US-09-922-261-81
3	1022	47.0	199	11	US-09-922-261-93
4	858	39.4	169	11	US-09-922-261-97
5	572	26.3	111	11	US-09-922-261-101
6	371	17.1	71	11	US-09-922-261-105
7	239	11.0	46	11	US-09-922-261-113
8	224	10.3	43	11	US-09-922-261-115
9	118	5.4	377	10	US-09-879-957-36
10	115	5.3	783	10	US-09-888-615-67
11	112	5.1	486	10	US-09-922-217-685
12	112	5.1	486	11	US-09-833-263-685
13	112	5.1	486	15	US-10-025-380-685
14	110.5	5.1	785	11	US-09-801-368-348
15	106	4.9	478	15	US-10-156-634A-2
16	106	4.9	599	11	US-09-758-269-6

17	106	4.9	1182	15	US-10-156-634A-4	Sequence 4, Appli
18	103.5	4.8	2224	15	US-10-115-563-14	Sequence 14, Appl
19	103.5	4.8	2224	15	US-10-172-712-31	Sequence 31, Appl
20	102.5	4.7	930	15	US-10-113-794A-1	Sequence 1, Appli
21	102.5	4.7	1243	15	US-10-196-935A-4	Sequence 4, Appli
22	101	4.6	504	15	US-10-029-180-6	Sequence 6, Appli
23	99.5	4.6	568	15	US-10-136-960-8	Sequence 8, Appli
24	99.5	4.6	663	12	US-09-932-257A-21	Sequence 21, Appli
25	99.5	4.6	730	12	US-09-932-257A-23	Sequence 23, Appli
26	99	4.6	656	10	US-09-825-301-9	Sequence 9, Appli
27	99	4.6	656	10	US-09-759-143-379	Sequence 379, App
28	99	4.6	656	10	US-09-780-669-379	Sequence 379, App
29	99	4.6	656	10	US-09-810-936-305	Sequence 305, App
30	99	4.6	656	10	US-09-822-827-379	Sequence 379, App
31	99	4.6	656	11	US-09-429-755-305	Sequence 305, App
32	99	4.6	656	11	US-09-924-400-305	Sequence 305, App
33	99	4.6	656	11	US-09-895-793-379	Sequence 379, App
34	99	4.6	656	11	US-09-895-814-379	Sequence 379, App
35	99	4.6	656	15	US-10-012-896-379	Sequence 379, App
36	99	4.6	656	15	US-10-010-940-379	Sequence 379, App
37	99	4.6	656	15	US-10-212-679-305	Sequence 305, App
38	99	4.6	864	11	US-09-883-096-2	Sequence 2, Appli
39	99	4.6	1719	10	US-09-759-143-378	Sequence 378, App
40	99	4.6	1719	10	US-09-780-669-378	Sequence 378, App
41	99	4.6	1719	10	US-09-822-827-378	Sequence 378, App
42	99	4.6	1719	11	US-09-895-793-378	Sequence 378, App
43	99	4.6	1719	11	US-09-895-814-378	Sequence 378, App
44	99	4.6	1719	15	US-10-012-896-378	Sequence 378, App
45	99	4.6	1719	15	US-10-010-940-378	Sequence 378, App

ALIGNMENTS

RESULT 1

US-09-922-261-77
; Sequence 77, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Purnanam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 77
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-77

Query Match	99.1%	Score	2155;	DB	11;	Length	412;
Best Local Similarity	100.0%	Pred. No.	6.9e-185;				
Matches	410;	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps	0;
QY	5	RWGEIPISSQTNRSSFDLLPREFLVVEHPDPLHQPSANKPKPTMLDIPSECSLTI	64				
Db	3	RWGEIPISSQTNRSSFDLLPREFLVVEHPDPLHQPSANKPKPTMLDIPSECSLTI	62				
QY	65	HTIQLIQRNRRLNLIATAQAQQOQTEGVKTESEPLSCPGSPPLPDDLPLDCKNPN	124				
Db	63	HTIQLIQRNRRLNLIATAQAQQOQTEGVKTESEPLSCPGSPPLPDDLPLDCKNPN	122				

```
QY 125 AFFQIRHSDPESDFYRGKEPVTLSWHSCROLLYQAVATILAHAGFDCANESVLETLTD 184
|
|
|
Db 123 AFFQIRHSDPESDFYRGKEPVTLSWHSCROLLYQAVATILAHAGFDCANESVLETLTD 182
|
|
|
QY 185 VAHEYCLKFTKLLRAVDREARLGOTPPDVMEQVFHEVGIGSVLSLQKFWOHRKIDYHS 244
|
|
|
Db 183 VAHEYCLKFTKLLRAVDREARLGOTPPDVMEQVFHEVGIGSVLSLQKFWOHRKIDYHS 242
|
|
|
QY 245 YMLQISKOLSEYERIVNPEKATEDAKPVKIKEEPPVSDITFPVSELEADLASGDOSLPM 304
|
|
|
Db 243 YMLQISKOLSEYERIVNPEKATEDAKPVKIKEEPPVSDITFPVSELEADLASGDOSLPM 302
|
|
|
QY 305 GVLGAQSRFFSNLEVEASPOASSAEVNASPLWNLAHVKMEPQSEEGNVSGHGVLSGDV 364
|
|
|
Db 303 GVLGAQSRFFSNLEVEASPOASSAEVNASPLWNLAHVKMEPQSEEGNVSGHGVLSGDV 362
|
|
|
QY 365 FEPMGSMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNORCKMRKI 414
|
|
|
Db 363 FEPMGSMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNORCKMRKI 412
|
|
|
RESULT 2
US-09-922-261-81
; Sequence 81, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 81
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-81
Query Match 87.3%; Score 1898; DB 11; Length 363;
Best Local Similarity 100.0%; Pred. No. 6.5e-162;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 52 MLDIPSECSLTITHTLIQHNRRLNLATAQAOQOQTEGVKTESEPLPSCGSPPL 111
|
|
|
Db 1 MLDIPSECSLTITHTLIQHNRRLNLATAQAOQOQTEGVKTESEPLPSCGSPPL 60
|
|
|
QY 112 PDDLPLDCKNPAPFOIRHSDPESDFYRGKEPVTLSWHSCROLLYQAVATILAHAGF 171
|
|
|
Db 61 PDDLPLDCKNPAPFOIRHSDPESDFYRGKEPVTLSWHSCROLLYQAVATILAHAGF 120
|
|
|
QY 172 DCANESVLETLTDVAHEYCLKFTKLLRAVDREARLGOTPPDVMEQVFHEVGIGSVLSL 231
|
|
|
Db 121 DCANESVLETLTDVAHEYCLKFTKLLRAVDREARLGOTPPDVMEQVFHEVGIGSVLSL 180
|
|
|
QY 232 QKFWOHRKIDYHSMLOISKOLSEYERIVNPEKATEDAKPVKIKEEPPVSDITFPVSEEL 291
|
|
|
Db 181 QKFWOHRKIDYHSMLOISKOLSEYERIVNPEKATEDAKPVKIKEEPPVSDITFPVSEEL 240
|
|
|
QY 292 EADLASGDOSLPMGVLGAQSRFFSNLEVEASPOASSAEVNASPLWNLAHVKMEPQSEEE 351
|
|
|
Db 241 EADLASGDOSLPMGVLGAQSRFFSNLEVEASPOASSAEVNASPLWNLAHVKMEPQSEEE 300
|
|
|
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 81
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-81
```

```
QY 352 GNVSGHGVLSGVDFEPMGSMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNORCKMR 411
|
|
|
Db 301 GNVSGHGVLSGVDFEPMGSMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNORCKMR 360
|
|
|
QY 412 RKI 414
|
|
|
Db 361 RKI 363
|
|
|
RESULT 3
US-09-922-261-93
; Sequence 93, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 93
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-93
Query Match 47.0%; Score 1022; DB 11; Length 199;
Best Local Similarity 100.0%; Pred. No. 1e-83;
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 216 MEQVFHEVGIGSVLSLQKFWOHRKIDYHSMLOISKOLSEYERIVNPEKATEDAKPVKI 275
|
|
|
Db 1 MEQVFHEVGIGSVLSLQKFWOHRKIDYHSMLOISKOLSEYERIVNPEKATEDAKPVKI 60
|
|
|
QY 276 KEEPVSDITFPVSELEADLASGDOSLPMGVLGAQSRFFSNLEVEASPOASSAEVNASP 335
|
|
|
Db 61 KEEPVSDITFPVSELEADLASGDOSLPMGVLGAQSRFFSNLEVEASPOASSAEVNASP 120
|
|
|
QY 336 LWNLAHVKMEPQSEEGNVSGHGVLSGVDFEPMGSMSEAGIPQSPDDSDSSYSGSHSTDS 395
|
|
|
Db 121 LWNLAHVKMEPQSEEGNVSGHGVLSGVDFEPMGSMSEAGIPQSPDDSDSSYSGSHSTDS 180
|
|
|
QY 396 LMGSSPVFNORCKMRKI 414
|
|
|
Db 181 LMGSSPVFNORCKMRKI 199
|
|
|
RESULT 4
US-09-922-261-97
; Sequence 97, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
```

```

; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 169
; TYPE: PRP
; ORGANISM: Homo sapiens
US-09-922-261-97

Query Match      39.4%; Score 858; DB 11; Length 169;
Best Local Similarity 100.0%; Pred. No. 4.1e-69;
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 MLIQISKQLSEYERIVNPKATEDAKPVKIKEEPVSDITFPVSEELADLASGDSPLMG 305
Db 1 MLIQISKQLSEYERIVNPKATEDAKPVKIKEEPVSDITFPVSEELADLASGDSPLMG 60

QY 306 VLGAQSERFPNLEVEASPOASSAEVNASPLNLAHVKNMEPQSEEGNVSGHVLGSDVF 365
Db 61 VLGAQSERFPNLEVEASPOASSAEVNASPLNLAHVKNMEPQSEEGNVSGHVLGSDVF 120

QY 366 EEPMSGMSEAGIPQSPDDSDSSYSGHSTDSLGMGSSPVFNQCKRMRKI 414
Db 121 EEPMSGMSEAGIPQSPDDSDSSYSGHSTDSLGMGSSPVFNQCKRMRKI 169

RESULT 5
US-09-922-261-101
; Sequence 101, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 111
; TYPE: PRP
; ORGANISM: Homo sapiens
US-09-922-261-101

Query Match      26.3%; Score 572; DB 11; Length 111;
Best Local Similarity 100.0%; Pred. No. 1e-43;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 MGVLGAQSERFPNLEVEASPOASSAEVNASPLNLAHVKNMEPQSEEGNVSGHVLGSD 363
Db 1 MGVLGAQSERFPNLEVEASPOASSAEVNASPLNLAHVKNMEPQSEEGNVSGHVLGSD 60

QY 364 VFEEPMGSMSEAGIPQSPDDSDSSYSGHSTDSLGMGSSPVFNQCKRMRKI 414
Db 61 VFEEPMGSMSEAGIPQSPDDSDSSYSGHSTDSLGMGSSPVFNQCKRMRKI 111

RESULT 6
US-09-922-261-105
; Sequence 105, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 105
; LENGTH: 71
; TYPE: PRP
; ORGANISM: Homo sapiens
US-09-922-261-105

Query Match      17.1%; Score 371; DB 11; Length 71;
Best Local Similarity 100.0%; Pred. No. 5.5e-26;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 344 MEPOESEGNGVSGHVLGSDVFEPEMSGMSEAGIPQSPDDSDSSYSGHSTDSLGMGSSPVF 403
Db 1 MEPOESEGNGVSGHVLGSDVFEPEMSGMSEAGIPQSPDDSDSSYSGHSTDSLGMGSSPVF 60

QY 404 NQCKRMRKI 414
Db 61 NQCKRMRKI 71

RESULT 7
US-09-922-261-113
; Sequence 113, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 113
; LENGTH: 46
; TYPE: PRP
; ORGANISM: Homo sapiens
US-09-922-261-113

Query Match      11.0%; Score 239; DB 11; Length 46;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 MSGMSEAGIPQSPDDSDSSYSGHSTDSLGMGSSPVFNQCKRMRKI 414
```

```
Db 1 MSGMSEAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQRCCKRMKI 46
|||||
RESULT 8
US-09-922-261-115
; Sequence 115, Application US/09922261
; Patent No. US2002011471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; PRIOR FILING DATE: 2001-08-03
; PRIOR FILING DATE: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-115

Query Match 10.3%; Score 224; DB 11; Length 43;
Best Local Similarity 100.0%; Pred. No. 4e-13;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 MSEAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQRCCKRMKI 414
|||||
Db 1 MSEAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQRCCKRMKI 43
|||||

RESULT 9
US-09-879-957-36
; Sequence 36, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; HOFFMAN, No. US20020034755A1h
; KAY, Brian K.
; FOWLKES, Dana M.
; MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
```

```
; FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 377 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-879-957-36

Query Match 5.4%; Score 118; DB 10; Length 377;
Best Local Similarity 21.0%; Pred. No. 0.034;
Matches 72; Conservative 39; Mismatches 108; Indels 124; Gaps 12;

QY 145 PVTLSWHSCROLLYQAVATILAHAGFDCANESVLETLTDV-----AHEYCLKFTKLLRF 199
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 5 PTVREAFH--KQMMGGFKETKEAEDGFRKAQKPAKLEVEAAKHAHAAKKEKL---- 59
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 200 AVDREAR-----LGOTPPPDVMEQ 218
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 60 AISREANSKADPSLNPEQLKKLDKIECKQDVLTKEKYEKSLUKELDQGTPOYMNMEQ 119
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 219 VFHEVGIGSVLSQKFWQHRIRKDYHMYLQISKQLS-----EYERIVNPERKAT 267
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 120 VFEQ-----CQFEEKRLRFFREVLLVLEQKHLDSLNVAGYKAIYHDLQESIRADAV 171
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 268 EDAPVKIKKEPVSDITFPVSEELADL-----ASGDQSLPMG 305
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 172 EDLRWFRANHGPGMANMNPQFEWSADLNRTLRSREKKKSTGVTLTGINTGTGQSLP-- 229
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 306 VLGAQSERFPNLEVEASPOASSAEVNASPLNLAHVKNMEPQSESEGVNKGVLGSDVF 365
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 230 -----SKPSTLNVPSNP-AQSAQSQS-----YNFDEDDT-----GSTVS 266
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 366 EEP-----MSGMSEAGIPQSPDDSDSYGSHSTDSLMGSSP 401
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 267 EKDDTKAKNVSSYEKTSYPTDWSDDSNPNFSTDANGDSNP 309
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 10
US-09-888-615-67
; Sequence 67, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 67
; LENGTH: 783
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-67
```

Query Match 5.3%; Score 115; DB 10; Length 783;
Best Local Similarity 21.1%; Pred. No. 0.19;
Matches 109; Conservative 61; Mismatches 178; Indels 168; Gaps 25;

QY 32 VEVDHPLP-SANKPKPTMLDIPSE-----PCSLTIHTIQLIQRRLRL 79
DB 1 MRVQDPTKALPEKAKSKRPT---VPHDESSDIAVGLTCQHVSHAISV-----NH 49

QY 80 IATAQONQOQTEGVTESEPLPCGSPPLPDDL-LPLDC-----KNPNAPFOIRHS 132
DB 50 VKRAIAEN---LWSVCSECLEERFYDGLVLTSDIWLCLKGFCGCKGNSQHSQSLKH- 105

QY 133 DPESDYRKGEK-----VTELSWHSCROLLYQAVATILAH----- 169
DB 106 -----FKSSRTEPHCIINLSTWIIWCYECDEKLSTHCNKKVLAQIVDFLQKHASKTQTS 160

QY 170 -----GDFCANESV-----LETLTDVAHE 198
DB 161 AFSRIMKLCEKEKETEIOKGGKCRNLRSVRGINLGNCTCFNNAVMQNLAQTYTLTDLMNE 220

QY 189 YCLKFTKLLRF-----AVDREARLGOTPPDVMEQVVFH---EVGIGSVLSLQKFWQ 236
DB 221 IKESSTKLTFFSSDQLOPLVLSRPG--PLTSALFLFLHSMKTEKGPLSPKVLFNQ 278

QY 237 -----HRIKDYHMYLQISKQL-----SEYERIV-----NPEKATED---AK 271
DB 279 LCQKAPRFKDFQO---QDSQELLHYLLDAVRTETRIQASILKAFNNPTTKTADDETRK 335

QY 272 PVKIK--EPPVSDITFPVSEE--LEADLASQDQSLPMGVLAGQASERFSPNLEVE--ASPQA 326
DB 336 KVAISTVKOPFIDISLPIIEERVSKPLLMGRNMKNYRSLRETDHRYSGNVNTIENIHOPRA 395

QY 327 SSAEVNASPLNLAHVK-----MEPQSEEGNVSNGHVLGSDVFEEMPMSGMSEA 375
DB 396 AKKHSSKDKSQLIHDRKCIKRLSSGETVYQKNENLNGDSLMFASLMN-----SES 449

QY 376 GIPQSPDDSSYSGSHSTDSL-MGSPVFNQRCRKR 410
DB 450 RLNESPTDDSEKASHESNVNADSPSESESAKQ 485

RESULT 11
US-09-922-217-685
; Sequence 685, Application US/0922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 685
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-685

Query Match 5.1%; Score 112; DB 10; Length 486;
Best Local Similarity 22.4%; Pred. No. 0.17;

Matches 56; Conservative 33; Mismatches 83; Indels 78; Gaps 9;

QY 192 KFTKLLRFAVDREARLGOTPPDVMEQVFEVIGSVLSLQKFWQHRKIDYHMYLQISK 251
DB 207 KYEKSLSL-----KELDOGTPOYMEMEQVFEQ-----CQOFEKRLRFRFVLLLEVQK 253

QY 252 QLS-----EYERIVNPEKATEDAKPVKIKEEVPVSDITFPVSELEADL----- 295
DB 254 HLNLNVAGYKAIYHDLQESIRAADAVEDLRFWRANHGPGMAMNWPQFEWSADLIRTLS 313

QY 296 -----ASDQSLPMGVLAGQASERFSPNLEVEASPOASSAEVNASPLWN 338
DB 314 RREKKATDGTFTLTGINOTGDQFLP-----SKPSSTLNVPSNP-AQSAQSS----- 360

QY 339 LAHVKNPEQSEEGNVSNGHVLGSDVFEEP-----MSGMSEAGIPQSPDDSDSYSGSH 391
DB 361 -----YNPEDEDDT-----GSTVSEKEDIKAKNVSSYEKTSQSYPTDWSDDSDSNPPS 408

QY 392 STDLSMGSSP 401
DB 409 STDANGDSNP 418

RESULT 12
US-09-833-263-685
; Sequence 685, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; CURRENT APPLICATION NUMBER: US/09/833,263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 685
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-833-263-685

Query Match 5.1%; Score 112; DB 11; Length 486;
Best Local Similarity 22.4%; Pred. No. 0.17;
Matches 56; Conservative 33; Mismatches 83; Indels 78; Gaps 9;

QY 192 KFTKLLRFAVDREARLGOTPPDVMEQVFEVIGSVLSLQKFWQHRKIDYHMYLQISK 251
DB 207 KYEKSLSL-----KELDOGTPOYMEMEQVFEQ-----CQOFEKRLRFRFVLLLEVQK 253

QY 252 QLS-----EYERIVNPEKATEDAKPVKIKEEVPVSDITFPVSELEADL----- 295
DB 254 HLNLNVAGYKAIYHDLQESIRAADAVEDLRFWRANHGPGMAMNWPQFEWSADLIRTLS 313

QY 296 -----ASDQSLPMGVLAGQASERFSPNLEVEASPOASSAEVNASPLWN 338
DB 314 RREKKATDGTFTLTGINOTGDQFLP-----SKPSSTLNVPSNP-AQSAQSS----- 360

QY 339 LAHVKNPEQSEEGNVSNGHVLGSDVFEEP-----MSGMSEAGIPQSPDDSDSYSGSH 391
DB 361 -----YNPEDEDDT-----GSTVSEKEDIKAKNVSSYEKTSQSYPTDWSDDSDSNPPS 408

QY 392 STDLSMGSSP 401
DB 409 STDANGDSNP 418

RESULT 13
US-10-025-380-685
; Sequence 685, Application US/10025380

Publication No. US20020182191A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yugu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clepper, Jonathan D.
APPLICANT: Skeiky, Yasir A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedwick, Thomas S.
APPLICANT: Carter, Darrick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C14
CURRENT APPLICATION NUMBER: US/10/025,380
CURRENT FILING DATE: 2001-12-19
NUMBER OF SEQ ID NOS: 1129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 685
LENGTH: 486
TYPE: PRT
ORGANISM: Homo sapiens
US-10-025-380-685

Query Match 5.1%; Score 112; DB 15; Length 486;
Best Local Similarity 22.4%; Pred. No. 0.17;
Matches 56; Conservative 33; Mismatches 83; Indels 78; Gaps 9;
QY 192 KFTKLLRAVDREARLGOTPPDWEQVFEHVGISVLSLQKFWQHRKIDYHSTMQLSK 251
DB 207 KYEKL-----KELDQGTPOYMNEMEQVFEQ-----COQFEKRLRFREVLLEVQK 253
QY 252 OLS-----EYERIVNPEKATEDAKPVKIKEPVSDDITFPVSEELADL----- 295
DB 254 HNLNSVAGYKAIYHDLQSTRAADAVEDLRFRANHGPGMAMNWPQFEWSADLIRTLS 313
QY 296 -----ASGQSLPMGVIGAQSERFPNSNLEVEASPOASSAEVNASPLWN 338
DB 314 RREKKATDGTLLTGINTGTGQFLP-----SKPSSTLNVSPN-AQSAQSQSS----- 360
QY 339 LAHYKMEQSEBEGNVSGHVLGSDVFEPP-----MSGMSEAGIQSPDDSDSSSYGSH 391
DB 361 -----YNPFEDEDDT-----GSTVSEKEDIKARNVSSYEKTSQYPTDWSDESNNPFS 408
QY 392 STDLSMGSSP 401
DB 409 STDANGDSNP 418

RESULT 14
US-09-801-368-348
Sequence 348, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Cali, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff

APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 348
LENGTH: 785
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-348

Query Match 5.1%; Score 110.5; DB 11; Length 785;
Best Local Similarity 18.8%; Pred. No. 0.48;
Matches 106; Conservative 67; Mismatches 189; Indels 201; Gaps 24;
QY 8 GEIPSSSOTNRSSFDL--LPREFRLVEVHPPLHQ-----PSANKPKPPTMLDI 55
DB 247 GSMSPNSNRTSRNSNSISSLAQPPMSNYQPSTYQYPGFKHTSSIPNSHSPIPRSLT 306
QY 56 PSE-----PCSLTIHTIQLIQRRLRLNLIAQAQNOQOQTEGVKTESEPLPSCPGS 108
DB 307 PTGPTSQNGPLSYNLQVGLL-----PQOQOQVSPLYDGNISITPPVKPST 353
QY 109 PP-----LPDILLPLDCKNPNAPFO---LRHSDPESDFYRGKE----- 144
DB 354 DQETYLTAHRHGVSDDQYDSMAKTMS--FQTTIRHPMLIATTNATGNSNTSGTSASIR 412
QY 145 -PVTLSWHSCROLLYQAVA---FILAHAGDFCANESVLETLTDVAHEYCLKFKLLRFA 200
DB 413 PRVTTTWEDEKTLCYQVEANGISVVRADNDWNGTKL-----LNVTKMTRGR 461
QY 201 VDREARLGOTPPDPMQVFEHVGISVLSLQKFW-----QHRKIDY----- 242
DB 462 RDGILK-----AEKIRHVVKIGS-MHLKGWIPFERALATAQREKIADYLVPLFIR 511
QY 243 --HSMYLOIS-----KOLSEE--YERIVN----- 262
DB 512 DIQSVLKQNNPNSDSSSSSTGIKISIPRTYQPINNYQNPNGPSNISAAQLTYSSMNL 571
QY 263 -----PEKATEDAKPVKIKEPVSDDITFPVSEELADLASGQSL-----PM--GVLG--A 309
DB 572 NKKIIPNNSIPAVSTIAAGERKPKKCTMPNSNQLEGHTITITLQTLSATMPMKQOLMGNIA 631
QY 310 QSERFPNSNLEVEASPOASSAEVNASPLWNLAHVKNMPEQSEEGNV----- 354
DB 632 SPLSYPRNATMNSASTLGITPADSKPL-TPSPTTTNTNQSSSNVGSVHTGITLPRVESE 690
QY 355 -----SCHGVLGSDVFEPPMSGM-----SEAGIQSPDD 383
DB 691 SASHSKWSKEADSGNTVPDNTQTLKEPRSSQLPISALTSTDTDKIKTSTSEATQPNPEPSE 750
QY 384 SDSSYGSHSTDLSM-GSSPVFNQ 405
DB 751 AEPVKESESQVDGAGDVSNE 773

RESULT 15
US-10-156-634A-2
Sequence 2, Application US/10156634A
Publication No. US20030013152A1
GENERAL INFORMATION:
APPLICANT: DeJong, Jeff L.
TITLE OF INVENTION: Transcription Factors Related to TFIIA
FILE REFERENCE: 119941-1092
CURRENT APPLICATION NUMBER: US/10/156,634A
CURRENT FILING DATE: 2002-05-28
PRIOR APPLICATION NUMBER: 09/326,529

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:55:28 ; Search time 31.3934 Seconds
(without alignments)
388.015 Million cell updates/sec

Title: US-09-857-308-1
Perfect score: 2175
Sequence: 1 MNLQRYWGEIPISSTQNRK.....SLMGSSPVFNQRCRMRKI 414

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2155	99.1	412	4	US-09-461-697-77
2	1898	87.3	363	4	US-09-461-697-81
3	1022	47.0	199	4	US-09-461-697-93
4	858	39.4	169	4	US-09-461-697-97
5	572	26.3	111	4	US-09-461-697-101
6	371	17.1	71	4	US-09-461-697-105
7	239	11.0	46	4	US-09-461-697-113
8	224	10.3	43	4	US-09-461-697-115
9	123	5.7	552	1	US-07-999-280A-22
10	123	5.7	552	1	US-08-426-279-22
11	123	5.7	552	1	US-08-401-013-22
12	123	5.7	552	3	US-08-426-570-22
13	123	5.7	552	3	US-08-425-876-22
14	123	5.7	552	4	US-08-426-243-22
15	123	5.7	552	4	US-08-401-632-22
16	118	5.4	377	4	US-08-630-915A-36
17	114	5.2	552	1	US-07-999-280A-24
18	114	5.2	552	1	US-08-426-279-24
19	114	5.2	552	1	US-08-401-013-24
20	114	5.2	552	3	US-08-426-570-24
21	114	5.2	552	3	US-08-425-876-24
22	114	5.2	552	4	US-08-426-243-24
23	114	5.2	552	4	US-08-401-632-24
24	106	4.9	478	4	US-09-326-529-2
25	106	4.9	1182	4	US-09-326-529-4
26	105	4.8	501	2	US-08-969-630-4
27	99	4.6	656	4	US-09-605-785-379

28	99	4.6	656	4	US-09-439-313-379	Sequence 379, App
29	99	4.6	656	4	US-09-352-616A-379	Sequence 379, App
30	99	4.6	1719	4	US-09-605-785-378	Sequence 378, App
31	99	4.6	1719	4	US-09-439-313-378	Sequence 378, App
32	99	4.6	1719	4	US-09-352-616A-378	Sequence 378, App
33	98.5	4.5	1342	1	US-07-978-895-4	Sequence 4, Appli
34	98.5	4.5	1342	2	US-08-473-119-4	Sequence 4, Appli
35	98.5	4.5	1342	2	US-08-475-352-4	Sequence 4, Appli
36	98.5	4.5	1343	6	5183884-4	Patent No. 5183884
37	98.5	4.5	1618	1	US-07-853-913-4	Sequence 4, Appli
38	98	4.5	822	4	US-08-684-932A-38	Sequence 38, Appl
39	97.5	4.5	1234	2	US-08-317-310A-15	Sequence 15, Appl
40	97.5	4.5	1234	5	PCT-US95-13041-15	Sequence 15, Appl
41	97.5	4.5	1243	2	US-08-557-139-2	Sequence 2, Appli
42	97.5	4.5	1342	2	US-08-484-438-9	Sequence 9, Appli
43	96.5	4.4	502	2	US-08-969-630-2	Sequence 2, Appli
44	96	4.4	1464	1	US-08-026-138E-1	Sequence 1, Appli
45	96	4.4	1706	2	US-08-459-568-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-461-697-77
Sequence 77, Application US/09461697
Patent No. 6277974
GENERAL INFORMATION:
APPLICANT: COGENT NEUROSCIENCE, Inc.
APPLICANT: Lo, Donald C.
APPLICANT: Barney, Shawn
APPLICANT: Thomas, Mary Beth
APPLICANT: Portbury, Stuart D.
APPLICANT: Puranam, Kasturi
APPLICANT: Katz, Lawrence C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
FILE REFERENCE: 10001-005-999
CURRENT APPLICATION NUMBER: US/09/461.697
CURRENT FILING DATE: 1999-12-14
NUMBER OF SEQ ID NOS: 466
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 77
LENGTH: 412
TYPE: PRT
ORGANISM: Homo sapiens
US-09-461-697-77

Query Match	99.1%	Score 2155;	DB 4;	Length 412;
Best Local Similarity	100.0%;	Pred. No. 4.9e+205;		
Matches 410;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	5	RYWGEIPISSTQNRSSFDLLPREFLVEVHDPLHPQPSANKPKPPTMLDIPSEPCSLTI	64	
Db	3	RYWGEIPISSTQNRSSFDLLPREFLVEVHDPLHPQPSANKPKPPTMLDIPSEPCSLTI	62	
Qy	65	HTIQLQHNRLRLNLIATAQAQNOQQTGKTESEPLPSCGSPPLPDDLLPLDCKNPN	124	
Db	63	HTIQLQHNRLRLNLIATAQAQNOQQTGKTESEPLPSCGSPPLPDDLLPLDCKNPN	122	
Qy	125	APQIRHSDPESFYRGKGPEVTVLSHWSHSCROLLYQAVATILAHAGFDCANESVLEFLTD	184	
Db	123	APQIRHSDPESFYRGKGPEVTVLSHWSHSCROLLYQAVATILAHAGFDCANESVLEFLTD	182	
Qy	185	VAHEYCLKFTKLLRFVAVDREARLGQTPFDVMEQVFEVGVIGSVLSQKFWQHRKIDYHS	244	
Db	183	VAHEYCLKFTKLLRFVAVDREARLGQTPFDVMEQVFEVGVIGSVLSQKFWQHRKIDYHS	242	
Qy	245	YMLQISKQLSEERYIVNPEKATEDAKPVKIKEPVSDDITFPVSEELADLASGDSLPM	304	
Db	243	YMLQISKQLSEERYIVNPEKATEDAKPVKIKEPVSDDITFPVSEELADLASGDSLPM	302	

QY 305 GVLGAQSRFPNLEVEASQAASSAEVNASPLWNLIAHYKMEPQSEEGNVSCHGVLSGV 364
Db 303 GVLGAQSRFPNLEVEASQAASSAEVNASPLWNLIAHYKMEPQSEEGNVSCHGVLSGV 362
QY 365 FEEPMSGSEAGIPQSPDDSSSYGSHSTDSLMGSSPVFNQCKRMKI 414
Db 363 FEEPMSGSEAGIPQSPDDSSSYGSHSTDSLMGSSPVFNQCKRMKI 412

RESULT 2

US-09-461-697-81
; Sequence 81, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 81
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-81

Query Match 87.3%; Score 1898; DB 4; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.1e-179;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 52 MLDIPSEPCSLTIHTIQLIQRNRLNLIATAQAOQOOTEVKTETSEPLPSCPGSPPL 111
Db 1 MLDIPSEPCSLTIHTIQLIQRNRLNLIATAQAOQOOTEVKTETSEPLPSCPGSPPL 60
QY 112 PDDLPLDCKNPAPFQIRHSDPESDFYRGKGEVPTLSWHSQRLLYQAVATILAHAGF 171
Db 61 PDDLPLDCKNPAPFQIRHSDPESDFYRGKGEVPTLSWHSQRLLYQAVATILAHAGF 120
QY 172 DCANESVLETLTDVAHEYCLFETKLLRPAVDREARLGOTPPDYMEQVFEVGVLSVL 231
Db 121 DCANESVLETLTDVAHEYCLFETKLLRPAVDREARLGOTPPDYMEQVFEVGVLSVL 180
QY 232 QKFWQHRKDYHSMQLQISKOLSEYERIVNPERKATEDAKPVKIKEEVPVSDITFPVSEEL 291
Db 181 QKFWQHRKDYHSMQLQISKOLSEYERIVNPERKATEDAKPVKIKEEVPVSDITFPVSEEL 240
QY 292 EADLASGQDQSLPMGVGAQSRFPNLEVEASQAASSAEVNASPLWNLIAHYKMEPQSEEE 351
Db 241 EADLASGQDQSLPMGVGAQSRFPNLEVEASQAASSAEVNASPLWNLIAHYKMEPQSEEE 300
QY 352 GNVSGHGVLSGVDVFEPMGSMSEAGIPQSPDDSSSYGSHSTDSLMGSSPVFNQCKRM 411
Db 301 GNVSGHGVLSGVDVFEPMGSMSEAGIPQSPDDSSSYGSHSTDSLMGSSPVFNQCKRM 360
QY 412 RKI 414
Db 361 RKI 363

RESULT 3

US-09-461-697-93
; Sequence 93, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:

; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 93
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-93

Query Match 47.0%; Score 1022; DB 4; Length 199;
Best Local Similarity 100.0%; Pred. No. 2.6e-93;
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 216 MEQVFHEVGIGSVLSLQKFWQHRKDYHSMQLQISKOLSEYERIVNPERKATEDAKPVKI 275
Db 1 MEQVFHEVGIGSVLSLQKFWQHRKDYHSMQLQISKOLSEYERIVNPERKATEDAKPVKI 60
QY 276 KEEPVSIDITFPVSEELADLASGQDQSLPMGVGAQSRFPNLEVEASQAASSAEVNASP 335
Db 61 KEEPVSIDITFPVSEELADLASGQDQSLPMGVGAQSRFPNLEVEASQAASSAEVNASP 120
QY 336 LWNLAHYKMEPQSEEGNVSCHGVLSGVDVFEPMGSMSEAGIPQSPDDSSSYGSHSTDS 395
Db 121 LWNLAHYKMEPQSEEGNVSCHGVLSGVDVFEPMGSMSEAGIPQSPDDSSSYGSHSTDS 180
QY 396 LMGSSPVFNQCKRMKI 414
Db 181 LMGSSPVFNQCKRMKI 199

RESULT 4

US-09-461-697-97
; Sequence 97, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-97

Query Match 39.4%; Score 858; DB 4; Length 169;
Best Local Similarity 100.0%; Pred. No. 3.5e-77;
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 MLQISKOLSEYERIVNPERKATEDAKPVKIKEEVPVSDITFPVSEELADLASGQDQSLPMG 305

```
Db 1 MLIQISQLSSEYERINVPKATEAKPVKIKEEPVSDITFPVSEELADLASGDSLPMG 60
QY 306 VLGAQSERFSPNLEVEASPOASSAEVNASPLWNLAHVKMEPQSEEGNVSGHVLGSDVF 365
Db 61 VLGAQSERFSPNLEVEASPOASSAEVNASPLWNLAHVKMEPQSEEGNVSGHVLGSDVF 120
QY 366 EEPMSGMSEAGIQSPDDSDSSYGSHSTDSLMGSSPVFNQCKKMRKI 414
Db 121 EEPMSGMSEAGIQSPDDSDSSYGSHSTDSLMGSSPVFNQCKKMRKI 169

RESULT 5
US-09-461-697-101
; Sequence 101, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-101

Query Match 26.3%; Score 572; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 3.8e-49;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 MGVLAGQSERFSPNLEVEASPOASSAEVNASPLWNLAHVKMEPQSEEGNVSGHVLGSD 363
Db 1 MGVLAGQSERFSPNLEVEASPOASSAEVNASPLWNLAHVKMEPQSEEGNVSGHVLGSD 60
QY 364 VFEPMMSGMSEAGIQSPDDSDSSYGSHSTDSLMGSSPVFNQCKKMRKI 414
Db 61 VFEPMMSGMSEAGIQSPDDSDSSYGSHSTDSLMGSSPVFNQCKKMRKI 111

RESULT 6
US-09-461-697-105
; Sequence 105, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 105
; LENGTH: 71
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-105

Query Match 17.1%; Score 371; DB 4; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.5e-29;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 344 MPEQSEEGNVSGHVLGSDVFEPMMSGMSEAGIQSPDDSDSSYGSHSTDSLMGSSPVF 403
Db 1 MPEQSEEGNVSGHVLGSDVFEPMMSGMSEAGIQSPDDSDSSYGSHSTDSLMGSSPVF 60

QY 404 NQCKKMRKI 414
Db 61 NQCKKMRKI 71

RESULT 7
US-09-461-697-113
; Sequence 113, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 113
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-113

Query Match 11.0%; Score 239; DB 4; Length 46;
Best Local Similarity 100.0%; Pred. No. 9e-17;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 MSGMSEAGIQSPDDSDSSYGSHSTDSLMGSSPVFNQCKKMRKI 414
Db 1 MSGMSEAGIQSPDDSDSSYGSHSTDSLMGSSPVFNQCKKMRKI 46

RESULT 8
US-09-461-697-115
; Sequence 115, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115
```



```

; INFORMATION FOR SEQ ID NO: 22:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 552 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-426-570-22

```

Query Match 5.7%; Score 123; DB 3; Length 552;
Best Local Similarity 21.6%; Pred. No. 0.0016;
Matches 90; Conservative 47; Mismatches 144; Indels 136; Gaps 20;

RESULT 13
US-08-425-876-22
; Sequence 22, Application US/08425876
; Patent No. 6117422
; GENERAL INFORMATION:
; APPLICANT: LADNER, MARTHA B.
; APPLICANT: NOBLE, JANELLE A.
; APPLICANT: MARTIN, GEORGE A.
; APPLICANT: KAWASAKI, ERNEST S.
; APPLICANT: COYNE, MAZIE YEE
; APPLICANT: HALENBECK, ROBERT F.
; APPLICANT: KOTHS, KIRSTON E.
; TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.

```

;
; REGISTRATION NUMBER: 31,395
; REFERENCE/DOCKET NUMBER: 0681.011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2718
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-425-876-22

Query Match          5.7%; Score 123; DB 3: Length 552;
Best Local Similarity 21.6%; Pred. No. 0.0016;
Matches 90; Conservative 47; Mismatches 144; Indels 136; Gaps 20;

QY 14 SSONTSSDPLLPRFLVEVDPLHOPSAKPKPTMLDIPSECSLIHTIIQLQHN 73
   | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 226 SQTEGSS--LLPSLPL-----RIEDPSAKQRP-----PRSTCQ-TLESTEQPNHG 270
   | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 74 RLRLNIATAQAQNOQQTGCVKTESEBPLSCGSPPLPDLLPLDCKNPAPFQIRHSD 133
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 271 DRL-----TEDSQPESGAGPVGVEDIL-----294

QY 134 PESDFYRGKEPVTELSSWISCRLLQQAVALIAHAGFDCANESVLETLTDVAHEYCLKF 193
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 295 -ESSL-GTNWVLEEASGEAEGFLAQ-----EAKFSPSPVGGSGS-----331

QY 194 TKLLREAFVDREARLGGTQPPF-----DVMQVDFHEVG-----IGSVLSLQKFWQHRIK 240
   : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 332 ---IQAETDRPALSAPPKSTQDKQPDVITDRPLTEVNPMPIGOT-----QNTPT 381

QY 241 DYHSYMLQISKQSEYERIVNPEKATDAKPKIKEEPVSDITFPVSELEADLASGDQ 300
   : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 382 E-----KTDGTSTLREHDQEPGSHIATPNQPVNSATPVAQLLLPKSH-----SWG 430

QY 301 SLPMGVL---GAQSERFPSNLVEASPOASS--AEVNASPLWNLAHVKMEPQSESEGN 353
   : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 431 VLPGLGEQKRTDRRRSFAELGGGSAEGAARPVARNFISIPITDTGHV-----EQHEG- 484

QY 354 VSGHGVGLGSDVFEEPMMSG-----MSRAGI-----PQSPDDSSSYGSHSTDLSL 396
   : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 485 -SSDPOIPRSVFHLLVPGITLVLITVGGILFYKWKWRSHRDPOTLDSSVGRPESSSL 540

```

RESULT 14
US-08-426-243-22
Sequence 22, Application US/08426243
Patent No. 6146851
GENERAL INFORMATION:
APPLICANT: LADNER, MARTHA B.
APPLICANT: NOBLE, JANELLE A.
APPLICANT: MARTIN, GEORGE A.
APPLICANT: KAWASAKI, ERNEST S.
APPLICANT: COYNE, MAZIE YEE
APPLICANT: HALENBECK, ROBERT F.
APPLICANT: KOTHS, KIRSTON E.
TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHIRON CORPORATION
STREET: Intellectual Property - R440, P.O. Box 8097
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0
Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/426,243
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/999,280
FILING DATE: 28-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: McGarrigle Jr., Philip L.
REGISTRATION NUMBER: 31,395
REFERENCE/DOCKET NUMBER: 0681.007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2718
TELEFAX: (510) 655-3542
TELEX: n/a
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 552 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-426-243-22

Query Match 5.7%; Score 123; DB 4; Length 552;
Best Local Similarity 21.6%; Pred. No. 0.0016;
Matches 90; Conservative 47; Mismatches 144; Indels 136; Gaps 20;
QY 14 SSOTNRSSFDLLPREFRLVEHDPPLHQPSSANKPKPTMLDIPSEPCSLTIHTIQLIQHN 73
DB 226 SORTEGSS--LLPSELPL-----RIEDPGSAKQRP-----PRSTCQ-TLESTEQPNHG 270
QY 74 RRLRLNIATAQAQNOQOQTEGVKTESEPLPFCGSPPLDPLDCKNPNAPFQIRHSD 133
DB 271 DRL-----TEDSQPHSAGGPVGVEDIL----- 294
QY 134 PSDFYRGKEPVTLSWHSQRLLYQAVATILAHAGFCANESVLETTDVAHEYCLKF 193
DB 295 -ESSL--GTNWLEEASGESEGLTQ-----EAKFSPSTPVGGS----- 331
QY 194 TKLLFAVDREARLGOTPPP-----DVMEQVFEVHG-----IGSVLSLQKFWOHRK 240
DB 332 ---IQAETDRPRALSAPPKSTEDQKPVDTIDRPLTEVNPMPICQT-----QNNTP 381
QY 241 DYHSYMLQISKOLSEYERIVNPEKATEDAKPVKIKEEPVSDITFPVSEELADLASGQ 300
DB 382 E---KTDGTSTLRDHPGSPHIATPNPQVNSATPVAQLLPKSH-----SWG 430
QY 301 SLPMGVL-----GAQSERFFSNLEVEASPOASS---AEVNASPLWNLAHVKMEPQSEEGN 353
DB 431 VLPLGELEGRSTRDRRSPAELEGGASGEAARPVARFNSIPLTDGTVH-----EQHEG- 484
QY 354 VSGHGVLSGDVFEPMGSG-----MSEAGI-----PQSPDSDSSYSGSHSTDLSL 396
DB 485 -SSDPOIPESVFHLLVPGIILVLLTVGGLLFYKWKWRSHRDPOTLDSVGRPEDSSL 540

RESULT 15

US-08-401-632-22
Sequence 22, Application US/08401632
Patent No. 6204020
GENERAL INFORMATION:
APPLICANT: LADNER, MARTHA B.
APPLICANT: NOBLE, JANELIE A.
APPLICANT: MARTIN, GEORGE A.
APPLICANT: KAWASAKI, ERNEST S.
APPLICANT: COYNE, MAZIE YEE
APPLICANT: HALENBECK, ROBERT F.
APPLICANT: KOTHS, KIRSTON E.
TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHIRON CORPORATION
STREET: Intellectual Property - R440, P.O. Box 8097

CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,632
FILING DATE: 09-MAR-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: McGarrigle Jr., Philip L.
REGISTRATION NUMBER: 31,395
REFERENCE/DOCKET NUMBER: 0681.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2718
TELEFAX: (510) 655-3542
TELEX: n/a
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 552 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-401-632-22

Query Match 5.7%; Score 123; DB 4; Length 552;
Best Local Similarity 21.6%; Pred. No. 0.0016;
Matches 90; Conservative 47; Mismatches 144; Indels 136; Gaps 20;
QY 14 SSOTNRSSFDLLPREFRLVEHDPPLHQPSSANKPKPTMLDIPSEPCSLTIHTIQLIQHN 73
DB 226 SORTEGSS--LLPSELPL-----RIEDPGSAKQRP-----PRSTCQ-TLESTEQPNHG 270
QY 74 RRLRLNIATAQAQNOQOQTEGVKTESEPLPFCGSPPLDPLDCKNPNAPFQIRHSD 133
DB 271 DRL-----TEDSQPHSAGGPVGVEDIL----- 294
QY 134 PSDFYRGKEPVTLSWHSQRLLYQAVATILAHAGFCANESVLETTDVAHEYCLKF 193
DB 295 -ESSL--GTNWLEEASGESEGLTQ-----EAKFSPSTPVGGS----- 331
QY 194 TKLLFAVDREARLGOTPPP-----DVMEQVFEVHG-----IGSVLSLQKFWOHRK 240
DB 332 ---IQAETDRPRALSAPPKSTEDQKPVDTIDRPLTEVNPMPICQT-----QNNTP 381
QY 241 DYHSYMLQISKOLSEYERIVNPEKATEDAKPVKIKEEPVSDITFPVSEELADLASGQ 300
DB 382 E---KTDGTSTLRDHPGSPHIATPNPQVNSATPVAQLLPKSH-----SWG 430
QY 301 SLPMGVL-----GAQSERFFSNLEVEASPOASS---AEVNASPLWNLAHVKMEPQSEEGN 353
DB 431 VLPLGELEGRSTRDRRSPAELEGGASGEAARPVARFNSIPLTDGTVH-----EQHEG- 484
QY 354 VSGHGVLSGDVFEPMGSG-----MSEAGI-----PQSPDSDSSYSGSHSTDLSL 396
DB 485 -SSDPOIPESVFHLLVPGIILVLLTVGGLLFYKWKWRSHRDPOTLDSVGRPEDSSL 540

Search completed: July 18, 2003, 10:01:22
Job time : 34.3934 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:03:57 ; Search time 25.5071 Seconds
(without alignments)
673.193 Million cell updates/sec

Title: US-09-857-308-1
Perfect score: 2175
Sequence: 1 MNLQRYWGEIPISSQTRNS.....SLMGSSPVFNQRCRMRKI 414

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	123	5.7	552	1 CSF1_MOUSE	P07141 mus musculus
2	117	5.4	486	1 PAC2_HUMAN	Q9unf0 homo sapien
3	117	5.4	575	1 SP6D_BACSU	P37963 bacillus su
4	112.5	5.2	1332	1 SP77_YEAST	P35177 saccharomyc
5	110.5	5.1	785	1 SOK2_YEAST	P53438 saccharomyc
6	109	5.0	486	1 PAC2_MOUSE	Q9wve8 mus musculus
7	108.5	5.0	671	1 CHS5_YEAST	Q12114 saccharomyc
8	106.5	4.9	1859	1 GBF1_HUMAN	Q92538 homo sapien
9	106	4.9	387	1 INVO_TARBA	P24711 tarsius ban
10	106	4.9	478	1 T2AY_HUMAN	Q9unn4 homo sapien
11	106	4.9	1182	1 T2AY_HUMAN	Q9y6g2 homo sapien
12	105.5	4.9	4377	1 ANK3_HUMAN	Q12955 homo sapien
13	105	4.8	488	1 PAC2_RAT	Q9qy17 rattus norv
14	105	4.8	896	1 Y249_HUMAN	Q92539 homo sapien
15	103.5	4.8	1324	1 SAL1_HUMAN	Q9nsc2 homo sapien
16	103.5	4.8	2224	1 FA5_HUMAN	P12259 homo sapien
17	103	4.7	830	1 VPP3_HUMAN	Q13488 h vacuolar
18	103	4.7	1462	1 NKCR_HUMAN	P30414 homo sapien
19	102.5	4.7	3358	1 PGCV_MOUSE	Q62059 mus musculus
20	102	4.7	1461	1 TOP2_CANAL	P87078 candida alb
21	102	4.7	3381	1 PGCV_BOVIN	P81282 bos taurus
22	101	4.6	317	1 APE_PAPAN	P05770 papio anubi
23	100.5	4.6	434	1 VATB_BORBU	O51120 borrelia bu
24	100.5	4.6	901	1 Y298_HUMAN	O15016 homo sapien
25	100	4.6	1153	1 PVD8_PLAKN	P50493 plasmodium
26	99.5	4.6	466	1 IMP2_DROME	P29681 drosophila
27	99.5	4.6	476	1 HLX1_MOUSE	Q61670 mus musculus
28	99.5	4.6	568	1 DCPY_ZYMO	P06672 zymomonas m
29	99.5	4.6	730	1 CCT2_HUMAN	O60583 homo sapien
30	99.5	4.6	1095	1 NEB1_RAT	O35867 rattus norv
31	99.5	4.6	1185	1 MAPX_DROME	P23226 drosophila
32	99.5	4.6	1235	1 IRS1_RAT	P35570 rattus norv
33	98.5	4.5	682	1 AMPH_CHICK	P50478 gallus gall

ALIGNMENTS				
RESULT 1				
ID	CSF1_MOUSE	STANDARD;	PRT;	552 AA.
AC	P07141;			
DT	01-APR-1988 (Rel. 07, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Macrophage colony stimulating factor-1 precursor (CSF-1) (MCSF).			
GN	CSF1 OR CSFM.			
OS	Mus musculus (Mouse), and			P21860 homo sapien
OS	Rattus norvegicus (Rat).			Q49429 mycoplasma
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			P48681 homo sapien
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			Q9nzm4 homo sapien
OX	NCBI_TaxID=10090, 10116;			Q9y6d6 homo sapien
RN	[1]			Q15911 homo sapien
RP	SEQUENCE FROM N.A.			Q26417 strongyloce
RC	SPECIES=Mouse;			Q14774 homo sapien
RX	MEDLINE=87174763; PubMed=3494232;			P35569 mus musculus
RA	Delamarter J.F., Hession C., Semon D., Gough N.M., Rothenbuhler R.,			P35568 homo sapien
RA	Mermoud J.-J.;			Q13796 homo sapien
RT	"Nucleotide sequence of a cDNA encoding murine CSF-1			
RL	(Macrophage-CSF).";			
RL	Nucleic Acids Res. 15:2389-2390(1987).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=Mouse;			
RX	MEDLINE=88320507; PubMed=2457916;			
RA	Ladner M.B., Martin G.A., Noble J.A., Wittman V.P., Warren M.K.,			
RA	McGrogan M., Stanley E.R.;			
RT	"cDNA cloning and expression of murine macrophage colony-stimulating			
RT	factor from L929 cells.;"			
RL	Proc. Natl. Acad. Sci. U.S.A. 85:6706-6710(1988).			
RN	[3]			
RP	SEQUENCE OF 1-100 FROM N.A.			
RC	SPECIES=Mouse;			
RX	MEDLINE=85242709; PubMed=3493488;			
RA	Rajavashisth T.B., Eng R., Shaddock R.K., Waheed A., Ben-Avram C.M.,			
RA	Shively J.E., Lusis A.J.;			
RT	"Cloning and tissue-specific expression of mouse macrophage colony-			
RT	stimulating factor mRNA.;"			
RL	Proc. Natl. Acad. Sci. U.S.A. 84:1157-1161(1987).			
RN	[4]			
RP	SEQUENCE OF 33-57.			
RC	SPECIES=Mouse;			
RX	MEDLINE=85242709; PubMed=3925458;			
RA	Ben-Avram C.M., Shively J.E., Shaddock R.K., Waheed A.,			
RA	Rajavashisth T.B., Lusis A.J.;			
RT	"Amino-terminal amino acid sequence of murine colony-stimulating			
RT	factor 1.;"			
RL	Proc. Natl. Acad. Sci. U.S.A. 82:4486-4489(1985).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=rat; STRAIN=Wistar; TISSUE=Muscle;			
RX	MEDLINE=93363632; PubMed=8357831;			
RA	Borycki A.G., Lenormund J., Guillier M., Leibovitch S.A.;			
RT	"Isolation and characterization of a cDNA clone encoding for rat			

CSF-1 gene. Post-transcriptional repression occurs in myogenic differentiation.";
 RL Biochim. Biophys. Acta 1174:143-152(1993).
 RN [6]
 RP SEQUENCE OF 1-13 FROM N.A.
 RC SPECIES-Mouse;
 RX MEDLINE=91340149; PubMed=1874443;
 RA Harrington M.A., Edenberg H.J., Saxman S.M., Pedigo L.M., Daub R.,
 RA Broxmeyer H.E.;
 RT Cloning and characterization of the murine promoter for the colony-stimulating factor-1-encoding gene.";
 RL Gene 102:165-170(1991).
 CC -!- FUNCTION: GRANULOCYTE/MACROPHAGE COLONY-STIMULATING FACTORS ARE CYTOKINES THAT ACT IN HEMATOPOIESIS BY CONTROLLING THE PRODUCTION, DIFFERENTIATION, AND FUNCTION OF 2 RELATED WHITE CELL POPULATIONS OF THE BLOOD. THE GRANULOCYTES AND THE MONOCYTES-MACROPHAGES.
 CC -!- FUNCTION: THIS CSF INDUCES MACROPHAGES
 CC -!- SUBUNIT: HOMODIMER, LINKED BY THREE DISULFIDE BONDS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X05010; CAA28660.1; -
 DR EMBL; M21952; AAA37481.1; -
 DR EMBL; M21149; AAA37482.1; -
 DR EMBL; M15692; AAA37480.1; -
 DR EMBL; M84361; AAA03032.1; -
 DR EMBL; M81316; AAA19866.1; -
 DR PIR; A23166; A23166.
 DR PIR; A26575; A26575.
 DR PIR; A31401; A31401.
 DR MGD; MGI:1339753; Csf1.
 KW Cytokine; Growth factor; Glycoprotein; Signal.
 FT SIGNAL 1 32
 FT CHAIN 33 552 MACROPHAGE COLONY STIMULATING FACTOR-1.
 FT DISULFID 63 63 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 39 122 BY SIMILARITY.
 FT DISULFID 80 171 BY SIMILARITY.
 FT DISULFID 134 178 BY SIMILARITY.
 FT DISULFID 189 189 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 191 191 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 292 292 D -> G.
 FT VARIANT 345 345 .S -> P.
 FT CONFLICT 3 3 MISSING (IN REF. 3).
 FT CONFLICT 6 6 A -> R (IN REF. 3).
 FT CONFLICT 7 8 AG -> PR (IN REF. 3).
 FT CONFLICT 246 246 P -> A (IN REF. 1).
 SQ SEQUENCE 552 AA; 60648 MW; 3886D72D70E770AF CRC64;
 Query Match 57%; Score 123; DB 1; Length 552;
 Best Local Similarity 21.6%; Pred. No. 0.31; 144; Indels 136; Gaps 20;
 Matches 90; Conservative 47; Mismatches 144; Indels 136; Gaps 20;
 QY 14 SSQTNRRSFDLLPREFRVLVEVDPLHPQSAKPKPTMLDIPSEPCSLTHTLIQHN 73
 DB 226 SORTEGSS--LLPSLPL-----RIEDPGSAKQRP-----PRSTCO-TLESTEQPNHG 270
 QY 74 RRLNLIATAQAQNOQOQTEGVKTESEPLPCSGPPPLDPLDCKNPNAPFOIRUSD 133
 DB 271 DRL-----TEDSQPHPSAGGVPVGVEDIL----- 294
 QY 134 PESDFYRGKGPVTELSWHSCROLLYQAVATILAHAGFDCANESVLETLTDAHEYCLKF 193
 DB 295 -ESSL--GTNVLEASGEASEGLFTQ-----EAKFSPTPVGGS----- 331

QY 194 TKLRFVAVDREARLQOTFPF-----DYMEQVFEHV-----IGSYLSLQKFWQHRK 240
 DB 332 --IQAETDRPRALSASPPKSTEDQKPVDTIDRPLTEVNMRIQOT-----QNNTPT 381
 QY 241 DYHSYMLQISKOLSEERYIVNPEKATEDARPKVKEEPVSDITFPVSEELADLASGDQ 300
 DB 382 E-----KTDGTSTLRDHQEGSPHIATPNQVNSATPFVAQLLLPKSH-----SWG 430
 QY 301 SLPMGLV----GAOSERFPNLEVEASPOASS---AEVNASPLNLAHVMEPOESERGN 353
 DB 431 VLPGLGELEKGRSTRDRRSPALEGSSAGEARPVARENSIPLTDTGCHV-----EQHEG- 484
 QY 354 VSGHVLGSDVFEPMVG-----MSEAGI-----PQSPDDSDSYGSHSDSL 396
 DB 485 -SSDPQIPESVFLHLLVPGIILVLTGVLGFLFYKWKRSRHPDQTLDSVGRPESSSL 540
 RESULT 2
 PAC2_HUMAN
 ID PAC2_HUMAN STANDARD; PRT; 486 AA.
 AC Q9UNF0; Q9H0D3; Q96HV9; Q9NPNI; Q95921; Q9Y4V2;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein kinase C and casein kinase substrate in neurons protein 2.
 GN PACSIN2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Retina, and Brain;
 EX MEDLINE=99358785; PubMed=10431838;
 RA Ritter B., Modregger J., Paulsson M., Plomann M.;
 RT "PACSIN 2, a novel member of the PACSIN family of cytoplasmic adapter proteins";
 RL FEBS Lett. 454:356-362(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Testis;
 RX MEDLINE=21154917; PubMed=11230166;
 RA Wienann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
 RA Ansgore W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
 RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
 RA Mexes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
 RA Wambutt R., Korn B., Klein M., Poustka A.;
 RT "Towards a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs";
 RL Genome Res. 11:422-435(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
 RA Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA Mcclay J., McLaren S., McMurray A., Milne S.A., Mortimore B.J.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,

RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Bemis G., Bentley J., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlffing T.,
 RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
 RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
 RA Kim O.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
 RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
 RA Wilkerson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
 RA Tilahun Y., Wright H.;
 RT "The DNA sequence of human chromosome 22.";
 RL Nature 402:489-495(1999).
 [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Skin;
 RA Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 [5]
 RN SEQUENCE OF 335-486 FROM N.A. (ISOFORM 2).
 RA Auffray C., Ansonge W., Ballabio A., Estivill X., Gibson K.,
 RA Lehrach H., Poustka A., Lundeberg J.;
 RT "The European IMAGE consortium for integrated molecular analysis of
 human gene transcripts.";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 [6]
 RN FUNCTION.
 RP PubMed=11082044;
 RA Modregger J., Ritter B., Witter B., Paulsson M., Plomann M.;
 RT "All three PACSIN isoforms bind to endocytic proteins and inhibit
 endocytosis.";
 RL J. Cell Sci. 113:4511-4521(2000).
 CC -|- FUNCTION: May play a role in vesicle formation and transport.
 CC -|- SUBUNIT: Homo- and hetero-aggregates with other PACSINS. Binds
 CC dynamin 1, synaptotagmin, synapsin 1 and the neural Wiskott-Aldrich
 CC syndrome protein (N-WASP) (By similarity).
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic. Vesicle-like cytoplasmic
 CC distribution (By similarity).
 CC -|- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
 CC produced by alternative splicing.
 CC -|- TISSUE SPECIFICITY: Ubiquitously expressed.
 CC -|- PTM: Phosphorylated by casein kinase 2 (CK2) and protein kinase C
 CC (PKC) (By similarity).
 CC -|- SIMILARITY: BELONGS TO THE PACSIN FAMILY.
 CC -|- SIMILARITY: CONTAINS 1 FCH DOMAIN.
 CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF128536; AAD41781.1; -;
 CC EMBL; AL136845; CAB66779.1; -;
 CC EMBL; AL022476; CAB39175.1; -;
 CC EMBL; AL049758; CAB51395.1; -;
 CC EMBL; BC008037; AAB08037.1; -;
 CC EMBL; AL389984; CAB97538.1; -;
 CC Genbank; HGNC:8571; PACSIN2.
 CC MIM; 604960; -;

DR HSP; Q60631; ICBQ.
 DR InterPro: IPR001060; Cdc15_Fes_CIP4.
 DR InterPro: IPR001452; SH3.
 DR Pfam; PF00611; FCH; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR PRODOM; PD000066; SH3; 1.
 DR SMART; SM00055; FCH; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS01333; FCH; 1.
 DR PROSITE; PS00002; SH3; 1.
 KW Endocytosis; SH3 domain; Coiled coil; Phosphorylation;
 KW Alternative splicing.
 FT DOMAIN 11 75 FCH.
 FT DOMAIN 426 486 SH3.
 FT DOMAIN 184 219 COILED COIL (POTENTIAL).
 FT VARIANTE 344 384 MISSING (IN ISOFORM 2).
 FT VARIANT 294 294 M -> I (IN DESNP:2746984).
 FT VARIANT 324 324 /FTID=VAR_013711.
 FT VARIANT 324 324 V -> F (IN DESNP:1062913).
 FT CONFLICT 182 182 L -> F (IN REF. 1).
 FT CONFLICT 256 256 D -> N (IN REF. 1).
 FT CONFLICT 309 309 N -> I (IN REF. 1).
 FT CONFLICT 336 336 S -> F (IN REF. 1).
 FT CONFLICT 378 380 DDT -> EDI (IN REF. 1).
 SQ SEQUENCE 486 AA; 55738 MW; 821DBEF65DAD1AA8 CRC64;
 Query Match 5.4%; Score 117; DB 1; Length 486;
 Best Local Similarity 22.8%; Pred. No. 0.67;
 Matches 57; Conservative 32; Mismatches 83; Indels 78; Gaps 9;
 QY 192 KFTKLLFAVDREARLQGPFPDVMEOVFHEVIGSVLSLQKFWOHRKIDYHVMLOISK 251
 DB 207 KYEKS-----KELDQGTPOYMENMEQVFQ-----CQFEKKLRFREVLLEVK 253
 QY 252 QLS-----EYERIVNPEKATEDAKPKIEPVSDITFPVSEELADL----- 295
 DB 254 HLDLSNVAGYKAIYHDLQESIRADAVEDLRFRANHGPGAMNWPQFEWSADLNRTLS 313
 QY 296 -----ASGQSLPMGVGLAQSERFSPNLEVASPOASAEVNASPLWN 338
 DB 314 RREKKKATDGVTLTGINTQTDGSLP-----SKPSTLNVPSPN-AQSAQSS----- 360
 QY 339 LAHVMEPOBSEEGNVSGHVLGSDVFEFP-----MSGMSEAGIPQSPDDSDSSVSGSH 391
 DB 361 -----YNPFEEDDT-----GSTVSEKDDTKAKNVSSYEKTSQTPDMSDDSENPF 408
 QY 392 STDLSMGSSP 401
 DB 409 STDANGSNP 418
 RESULT 3
 ID SP6D_BACSU STANDARD; PRT; 575 AA.
 AC P37963;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Stage VI sporulation protein D.
 GN SPOVID.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP MEDLINE=93194796; PubMed=8449878;
 RA Beall B., Driks A., Losick R., Moran C.P. Jr.;
 RT "Cloning and characterization of a gene required for assembly of the
 Bacillus subtilis spore coat.";
 RL J. Bacteriol. 175:1705-1716(1993).
 RN [2]

```

RP SEQUENCE FROM N.A.
RX MEDLINE=98044033; PubMed=9384377;
RA Azevedo V., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillett S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Evington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinis S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogilwa A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Soldo B.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpatra P., Togononi A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zunshein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -!- FUNCTION: REQUIRED FOR ASSEMBLY OF A NORMAL SPORE COAT. MAY BE A
CC COMPONENT OF THE INNERMOST LAYER OF THE SPORE COAT THAT AIDS IN
CC ITS ADHERENCE TO THE PRESPORE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSION IS INITIATED AROUND THE SECOND
CC HOUR OF SPOREALATION AND CONTINUES THROUGHOUT DEVELOPMENT. MAY BE
CC EXPRESSED PREDOMINANTLY IN THE MOTHER CELL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L07792; AAA22808.1; -;
DR EMBL; Z99118; CAB14771.1; -;
DR Subtilist; BG10346; spoVID.
DR InterPro; IPR002482; LysM.
DR Pfam; PF01476; LysM; 1.
DR SMART; SM00257; LysM; 1.
KW Sporulation; Complete proteome.
SQ SEQUENCE 575 AA; 64976 MW; 9A879AB16B18884F CRC64;
Query Match 5.4%; Score 117; DB 1; Length 575;
Best Local Similarity 21.4%; Pred. No. 0.84; Indels 108; Gaps 20;
Matches 87; Conservative 53; Mismatches 158;
QY 9 EIPISSTQNRSSDLLPRFLVEVHDPPLH-----QFSANKPKPTML-DIPSEPCSLT 63
DQ III: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 146 EIPLYEA-----PAAPREELSEPPAHSVVEEPGASSAEAVLQHEPPAEPPEL- 194
QY 64 IHTQLQHNRLRLNLATAQAOQOQTEGVKTESEPLSCPSPLPDDL----- 115
DQ III: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 195 -----FISKAGLRELETE-----KAESPPESVASEPEAREDKVEERSEE 236
QY 116 LPLDCKNPNAPFQIRHSDPED-----FYRCK---GEPTVLSWHSCROLLYQAVAT 164
DQ I: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 237 LAVPETEVRASETESEPEPDSPSEIETQIVAKKETAEPAIAA--DVREEDADSPAET 294

```

```

QY 165 IL-AHAG-----FDCANESVL-----EFLTDAVAHEYCLKFTKLLRFAVDREARLGQT 210
DQ I: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 295 ELREHVGAEESPAELAEHLSETVIKAKEETVSPNHEYALR-----QEAQNEEAASDQ 349
QY 211 PFDV-MEOPVHEVGISV-----LSLOK-----FWOHRKIDYHSMVQ-----IS 250
DQ I: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 350 ADPALCOEEAPDEALESVSAALSIEDSRETASAVYMENDNADLHFHFNKTSSEASQ 409
QY 251 KQLSEEEYRIVNPEKATED---AKPVKIKPEVSDITFPVSELEADLASDQSLPMGVL 307
DQ I: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 410 EELPEPAYRFLPQEEEDSFYSAPKLEEEQEESFEIEVRKTPSAEEPKETPF--- 466
QY 308 GAQSERPPSNLEVASPAQSAEYVNASPLMLAHVKMPEQSESGN 353
DQ I: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 467 --QSFLPESSETERK-----ETDAVPRVAPAAETKEPQTKESDN 504

```

RESULT 4

ID	SPT7_YEAST	STANDARD	PRT	1332 AA
AC	P35177			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Transcriptional activator SPT7.			
GN	SPT7 OR YBR081C OR YBR0739.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C;			
RC	MEDLINE=95229044; PubMed=7713415;			
RA	Gansheroff L.J., Dollard C., Tan P., Winston F.;			
RT	"The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein			
RT	important for transcription in vivo.";			
RL	Genetics 139:523-536(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C;			
RC	MEDLINE=95076715; PubMed=7985423;			
RA	van der Aart Q.J.M., Barthe C., Dolignon F., Aigle M., Crouzet M.,			
RA	Steenma H.Y.;			
RT	"Sequence analysis of a 31 kb DNA fragment from the right arm of			
RT	Saccharomyces cerevisiae chromosome II.";			
RL	Yeast 10:959-964(1994).			
RN	[3]			
RP	SEQUENCE OF 1-835 FROM N.A.			
RC	STRAIN=S288C;			
RC	MEDLINE=92285152; PubMed=1350857;			
RA	Haynes S.R., Dollard C., Winston F., Beck S., Trowsdale J.,			
RA	David I.B.;			
RT	"The bromodomain: a conserved sequence found in human, Drosophila and			
RT	yeast proteins.";			
RL	Nucleic Acids Res. 20:2603-2603(1992).			
CC	-!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY			
CC	OTHER GENES.			
CC	-!- SUBCELLULAR LOCATION: Nuclear.			
CC	-!- SIMILARITY: CONTAINS 1 BROMODOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			

```
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L22537; AAC37424.1; -
CC DR EMBL: X76294; CAA53940.1; -
CC DR EMBL: Z35950; CAA85026.1; -
CC DR EMBL: M87651; AAA35087.1; -
CC DR PIR: S41552; S41552.
CC DR HSSP: Q92831; I891.
CC DR TRANSFAC: T04835; -
CC DR SGD: S000285; SPW7.
CC DR InterPro: IPR001487; Bromodomain.
CC DR Pfam: PF00439; bromodomain; 1.
CC DR PRINTS: PR00503; BROMODOMAIN.
CC DR SMART: SM00297; BROMO; 1.
CC DR PROSITE: PS00633; BROMODOMAIN_1; 1.
CC DR PROSITE: PS50014; BROMODOMAIN_2; 1.
CC DR Transcription regulation; Nuclear protein; Activator; Bromodomain.
CC FT DOMAIN 458 528 BROMODOMAIN.
CC SQ SEQUENCE 1332 AA; 152616 MW; 083B63624669244F CRC64;

Query Match
Best Local Similarity 5.2%; Score 112.5; DB 1; Length 1332;
Matches 62; Conservative 49; Mismatches 96; Indels 83; Gaps 13;

CC QY 122 NPNAFQIRK-----SDPESDFYRGKGPVTVLSWHSCR-----QLLYQAVATILAH 168
CC Db 940 NPNA-FLNNRIYNTIIDSLDI-----DPVSQPLTHDYKNNRELIIWKPMHKNISKVAMA 993
CC QY 169 AGFDCANESVLETLTDVAHEYCLKFKTLRFVADREARLGQTFPDVMEQVHEVGIGSV 228
CC Db 994 NGFETAHPSAINMLTEIAGDYLNLIKTLKHETNS-LNRGTNVEMLQTTLLENGI--- 1049
CC QY 229 LSLQRFWQHRIRKIDYHSY-----LQISKQLSEYER-IVNP-----EKATEDAK 271
CC Db 1050 -----NRPDLSFYVSEFGKTKKLDIKQKLESFLRALLRPTQLSENFEDS 1101
CC QY 272 PVKIKEPVSDIT-----FPVSE-ELEADLASGDQLPMGVGAQSERFPNSLEVASPOA 326
CC Db 1102 QSFFTGDFASELTGDFGFRLEGLKEFGVLSVPLQLLTQTQTVDGETKVQAK--- 1158
CC QY 327 SSAEVNASPLNLAHVHKMPEQSEEGNSGCHGVLSGDVFEPEPMGNSGAG 376
CC Db 1159 -----KIQPESD-----SIVYKTKTKGMLDAG 1181

RESULT 5
ID: SOK2_YEAST STANDARD; PRT; 785 AA.
AC P53438;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SOK2 protein
GN SOK2 OR YMR016C OR YMR711.03C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96069395; PubMed=8524252;
RA Ward M.P., Gimeno C.J., Fink G.R., Garrett S.;
RA Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PLAYS A GENERAL REGULATORY ROLE IN THE CYCLIC AMP-
CC DEPENDENT PROTEIN KINASE-STIMULATED (PKA) SIGNAL TRANSDUCTION
CC PATHWAY BY REGULATING THE EXPRESSION OF GENES IMPORTANT IN GROWTH
```

```
CC AND DEVELOPMENT. MAY INHIBIT THE SWITCH FROM UNICELLULAR TO
CC FILAMENTOUS GROWTH.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: BELONGS TO THE EFG1/PHD1/STUA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S80522; AAB35749.1; -
CC DR EMBL: Z49211; CAA89117.1; -
CC DR TRANSFAC: T03616; -
CC DR SGD: S0004618; SOK2.
CC DR InterPro: IPR003163; Yeast_DNA_bind.
CC DR Pfam: PF02292; Yeast_DNA_bind; 1.
CC DR Transcription regulation; Nuclear protein; DNA-binding.
CC FT DOMAIN 38 45 POLY-GLN.
CC FT DOMAIN 148 151 POLY-TYR.
CC FT DOMAIN 169 172 POLY-TYR.
CC FT DOMAIN 206 217 POLY-GLN.
CC FT DOMAIN 331 335 POLY-GLN.
CC FT DOMAIN 526 532 POLY-SER.
CC SQ SEQUENCE 785 AA; 85643 MW; F4E0F224BBE84645F CRC64;

Query Match
Best Local Similarity 5.1%; Score 110.5; DB 1; Length 785;
Matches 106; Conservative 67; Mismatches 189; Indels 201; Gaps 24;

CC QY 8 GEIPTSSTNRSSFDL--LPREFLVEVHDPLHQ-----PSANKPKPTMLDI 55
CC Db 247 GMSPNSTNSRNSNSISSLAQPPMSNTPQSTYQPGFHTSSIPNSHSIPPRSLIT 306
CC QY 56 PSE-----PCSLTIHTLIQHNRRLNLATAQAQQOQTEGVKTESEPLPSCPS 108
CC Db 307 PQGGTSSQNGPLSYNLQVGLL-----PPQQQQQVSPLYDGNSTTPPVKPT 353
CC QY 109 PP-----LPDDLPLDCKNPNAPQ---IRHSDPESDFYRGKE----- 144
CC Db 354 DQETVLTANRHGVSDQYDSMAKTWNS-FQTTIRHPMLIATTNATGNTSOTSASIR 412
CC QY 145 -PVTELSWHSCROLLYQAVA---TILAHAGFCANESVLETLTDVAHEYCLKFKTLRFA 200
CC Db 413 PRVTMTWDEKTLCYQVEANGISVVRADNDWNGTKL-----LNVTKTRGR 461
CC QY 201 VDREARLGQTFPDVMEQVHEVGIGSVLSLQKF-----QHRIKDY----- 242
CC Db 462 RDGILK-----AEKIRHVVKIGS-MHLKGWIPFERALAIQREKIADYLYPLFIR 511
CC QY 243 --HSYMLQIS-----KOLSEE--YERIVN----- 262
CC Db 512 DIQSVLKQNNPNDSSSSSSTGKISPRITYQPINNYQNPNGPSNISAQOITYSSMNL 571
CC QY 263 -----PEKATEDAKPVKIKEPVSDITFPVSELEADLASGDQL-----PM--GVLG--A 309
CC Db 572 NKKIIPNNSIPAVSITIAAGEKPLKCTMPNSQLEGHTITNQLTSLATMPKQQLMGNTA 631
CC QY 310 QSERFPNSLEVASPOASAEVNASPLNLAHVHKMPEQSEEGNV----- 354
CC Db 632 SPLSPYRNATMNSASTLGTITPADSKPL-TPSPTTTTNQSSSNVSGSIHTGTLPRVSE 690
CC QY 355 -----SGHCVLSGDVFEPEPMSCM-----SEAGIPQSPDD 383
CC Db 691 SASHSKWSKEADSGNTVDPDNTLKEPRSQLPISALTSTDTDKIKTSTDEATQNPPESE 750
CC QY 384 SDSSYSGSHSTDSLM-GSSSPVFNQ 405
CC Db 751 AEPVKESESKSQVDGAGDVSNE 773
```

RESULT 6
 PAC2_MOUSE
 ID PAC2_MOUSE STANDARD; PRT; 486 AA.
 AC Q9WVE8;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein kinase C and casein kinase substrate in neurons protein 2.
 GN PACSIN2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 x DBA;
 RX MEDLINE=99358785; PubMed=10431838;
 RA Ritter B., Modregger J., Paulsson M., Plomann M.;
 RT "PACSIN 2, a novel member of the PACSIN family of cytoplasmic adapter
 proteins.";
 RL FEBS Lett. 454:356-362(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Strausberg R.;
 RT Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP FUNCTION.
 RX PubMed=11082044;
 RA Modregger J., Ritter B., Witter B., Paulsson M., Plomann M.;
 RT "All three PACSIN isoforms bind to endocytic proteins and inhibit
 endocytosis.";
 RL J. Cell Sci. 113:4511-4521(2000).
 CC -!- FUNCTION: May play a role in vesicle formation and transport.
 CC -!- SUBUNIT: Homo- and hetero-aggregates with other PACSINs. Binds
 CC dynamin 1, synaptotagmin, synapsin 1 and the neural Wiskott-Aldrich
 CC syndrome protein (N-WASP) (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Vesicle-like cytoplasmic
 CC distribution (By similarity).
 CC -!- TISSUE SPECIFICITY: Ubiquitously expressed, with highest levels in
 CC brain, heart, skeletal muscle and ovaries.
 CC -!- PH: Phosphorylated by casein kinase 2 (CK2) and protein kinase C
 CC (PKC) (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE PACSIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 FCH DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF128535; AAD41780.1; -;
 DR EMBL; BC023502; AAH23502.1; -;
 DR HSSP; P29355; 1SEM.
 DR MGD; MGI:1345153; Pacsin2.
 DR InterPro; IPR001060; Cdc15_Fes_CIP4.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00611; FCH; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR PRODOM; PD000066; SH3; 1.
 DR SMART; SM00055; FCH; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS0133; FCH; 1.
 DR PROSITE; PS00002; SH3; 1.
 DR Endocytosis; SH3 domain; Coiled coil; Phosphorylation.
 FT DOMAIN 11 75 FCH.
 FT DOMAIN 426 486 SH3.
 FT DOMAIN 184 239 COILED COIL (POTENTIAL).

SQ SEQUENCE 486 AA; 55833 MW; 66C17ECC6767E0E7 CRC64;
 Query Match 5.0%; Score 109; DB 1; Length 486;
 Best Local Similarity 21.4%; Pred. No. 2.4;
 Matches 69; Conservative 35; Mismatches 94; Indels 124; Gaps 13;
 Oy 167 AHAGFDGANEVLETLTDV-----AHEYCLKFTKLLRFAVDREA----- 205
 Db 134 AEDGFRKAQKPAWKKLKEVEAAKHAHTACKEEKL---AISREANSKADPSLNPEQLKKL 190
 Oy 206 -----RLGQT--PFPPDMEQVFHEVGIGSVLSLQKFWQHRIK 240
 Db 191 QDKTEKCKQDVLTKDKYKSLKELDTTPQYMEQMEOVFEEQ-----CQFEKKLR 242
 Oy 241 DYHSYMLQISKOLS-----EEYRTVNPKEKATEDAKPVKIEEPVSDITFPVSE 289
 Db 243 FFREVLLVQKHLDLSNVASYKTYIRELEQSIKAADAVEDLRFRANHGPGMANWQFQE 302
 Oy 290 ELEADL-----ASGDQLPMGVGLGAQSERFP-SNLEVEASQA 326
 Db 303 EWSADLNLRLSRREKKKAVDGVTLTGINTGDQS-----GQNKPGSNLSVPSNP-A 352
 Oy 327 SSAEVNASPLNLAHVKNPEQSEEGNVSGHVLGSDVFEFP-----MSGMSEAGIPQ 379
 Db 353 QSTQLQSS-----YNPFEDEDT-----GSSISEKEDIKAKNVSSYEKTQTYPT 396
 Oy 380 SPDDSDSYSGSHSTDSLGMSSP 401
 Db 397 DWSDDSENPFSSDANGDSNP 418

RESULT 7

CHS5_YEAST
 ID CHS5_YEAST STANDARD; PRT; 671 AA.
 AC Q12114;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Chitin biosynthesis protein CHS5 (CAL3 protein).
 GN CHS5 OR CAL3 OR YLR330W OR L8543.18.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GRF88;
 RX MEDLINE=97265380; PubMed=9111317;
 RA Santos B., Duran A., Valdivieso M.H.;
 RT "CHS5, a gene involved in chitin synthesis and mating in
 RT Saccharomyces cerevisiae.";
 RL Mol. Cell. Biol. 17:2485-2496(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favellio A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
 RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
 RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
 RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
 RA Taich A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
 RA Wilson R., Waterston R.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP FIBRONECTIN TYPE III DOMAIN.
 RX MEDLINE=97148176; PubMed=8994808;
 RA Bateman A., Chothia C.;
 RT "Fibronectin type III domains in yeast detected by a hidden Markov
 RT model.";
 RL Curr. Biol. 6:1544-1546(1996).
 CC -!- FUNCTION: INVOLVED IN CHITIN SYNTHESIS AND ALSO REQUIRED FOR
 CC MATING.
 CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.


```
RESULT 9
INVO_TARBA
ID INVO_TARBA STANDARD; PRT: 387 AA.
AC P24711;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Involutrin.
GN IVL.
OS Tarsius bancanus (Western tarsier) (Horsfield's tarsier).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Tarsi; Tarsiidae; Tarsius.
OX NCBI_TaxID=9477;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91271381; PubMed=1905021;
RA Djian P., Green H.;
RT "Involutrin gene of tarsoids and other primates: alternatives in
  evolution of the segment of repeats.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:5321-5325(1991);
CC -!- FUNCTION: Part of the insoluble cornified cell envelope (CE) of
  stratified squamous epithelia.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Constituent of the scaffolding
  of the cornified envelope.
CC -!- TISSUE SPECIFICITY: Keratinocytes of epidermis and other
  stratified squamous epithelia.
CC -!- PM: Substrate of transglutaminase. Specific glutamines or lysines
  are cross-linked to keratins, desmoplakin and to inter involucrin
  molecules.
CC -!- SIMILARITY: BELONGS TO THE INVOLUCRIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M5124; AAA36960.1; -
CC PIR; A43704; A43704.
CC InterPro; IPR002360; Involutrin.
CC PROSITE; PS00795; INVOLUCRIN; 1.
CC Keratinization; Repeat.
KW SEQUENCE 387 AA; 45077 MW; 146BAC2E490AA805 CRC64;
Query Match 4.9%; Score 106; DB 1; Length 387;
Best Local Similarity 23.0%; Pred. No. 2.8;
Matches 76; Conservative 37; Mismatches 140; Indels 78; Gaps 17;
QY 77 RNLATAQAQNOQQTGKTESEPLSCPGSPPLPDDLLPLDCKN-PN---AFQIRHS 132
DB 17 QELLKTVPPANTQODQKQTPSPAPCKGSPSELVPEKHPAPYKVPQEQCEPQQDQH 76
QY 133 DPESDFVRGKEPVTLSWHSRCLLYQAVATILAHAGFCANESVLETLTDVAHEYCLK 192
DB 77 EPELQLGRKQEP-QEQEVHPGKQ--QKQPEQEAHLG-----KKQEPQEQEVH----- 122
QY 193 FTKLLRFADVAREALRGTPFPDVMQVFEHVGIGSVLSLQKFWQHRKIDYHMYLQISKQ 252
DB 123 LGKQKQKTEQEVHLGKQK-QELQEQEVH-----LEKQLQEP-QEVH-----LEKQ 166
QY 253 LSEYERIVNPEKATEAKPV-----KIKEPVSDITFPVSELEADLASGDSQSLPMGV-- 306
DB 167 LQEQE---VHLEKQLQEPPELNLGKQEQEP-----QEQEAYLGRKQEQELPEPQDP 214
QY 307 ---LCAQSERPFSNLEVEASQAASSAEVNASPLWNLAVHVKMEPQSEEGNVSCHGLGSD 363
DB 215 ELHLGKQKQEQE-PQEQEVLEKQEQEQEQE-----LHLGKQEQEQE----- 253
QY 364 VFPEPMSCMGSEAGIPQSPDDSDSSSYGSHSTD 394
DB 254 --EQELHLRKLQOVPOEPQDQELHLGKQEQE 282
RESULT 10
T2AV_HUMAN
ID T2AV_HUMAN STANDARD; PRT: 478 AA.
AC O9UN4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE TFIIA-alpha and beta like factor (ALF).
GN ALF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99292779; PubMed=10364255;
RA Upadhyaya A.B., Lee S.H., Dejong J.;
RT "Identification of a general transcription factor TFIIA(alpha)/(beta)
  homolog selectively expressed in testis.";
RL J. Biol. Chem. 274:18040-18048(1999).
CC -!- FUNCTION: MAY FUNCTION AS A TESTIS SPECIFIC TRANSCRIPTION FACTOR.
  BINDS DNA IN CONJUNCTION WITH GTF2A2 AND TBP (THE TATA-BINDING
  PROTEIN) AND TOGETHER WITH GTF2A2, ALLOWS MRNA TRANSCRIPTION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Testis-specific.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF106857; AAD39634.1; -
CC HSP; P32773; IYTF.
CC TRANSFAC; T05004; -
CC MIN; 605358; -
CC InterPro; IPR004855; TFIIA.
CC Pfam; PF03153; TFIIA; 1.
CC Transcription regulation; Nuclear protein; DNA-binding.
KW SEQUENCE 478 AA; 52445 MW; 0EF012A79CD17619 CRC64;
Query Match 4.9%; Score 106; DB 1; Length 478;
Best Local Similarity 22.9%; Pred. No. 3.8;
Matches 72; Conservative 42; Mismatches 116; Indels 84; Gaps 15;
QY 128 QIRHSDPESDFYRK-GEPTVTELSW-HSCROLLYQAVATILAHAG-----FDCAN 175
DB 46 KVLQSKATEDFFRNSIOSPLFTLQPLHSLHQTLOSSTASLVPAGRTLPSTTAEGLTGN 105
QY 176 ES-----VLETLTDVAHEYCLKTKLIRFAVDREARLGQTFPDVMEQV 219
DB 106 SSANFTFGPIPIHVPAGVTLOTVS--GHLYKNVPIWV---TETSGRAGILQHP--IQOV 158
QY 220 FHEVGIGSVLSLQKFWQHRKIDYHMYLQISKQISEEYERIVNPEKATEAKPVKIK--E 277
DB 159 FQQLGQPSVI-----QTSVPQLNPWSLQATTEKRSQRTETVLQ-QPAILPSGPDVDRKHL 211
QY 278 EPVSDITFPVSEE-----LEADLASGQ-----SLPMGVLAGQSERFSPNLEVEASQAASSA 329
DB 212 NATSDILVSPGNEHKIYVPEALLCHQESSHYISLPGVVFSPQVSTNSDVSLSGSASMA 271
QY 330 EVNASPLWNLAVHVKMEPQSEEGNVSCHGLV-----GSDVFEEPMGSM 372
DB 272 Q-----NL-----HDESLSYPRGALHQHVTDIOLHLKRMVGCDSVKQPRNIE 316
```



```
QY 373 SEAGIQSPDDSDS 386
Db 317 EPSNIPVSEKDSNS 330

RESULT 11
T2AZ_HUMAN STANDARD; PRT; 1182 AA.
AC Q9Y6Q2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Stoned B-TFIIA-alpha and beta like factor (SALF).
GN SALF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99292779; PubMed=10364255;
RA Upadhyaya A.B., Lee S.H., DeJong J.;
RT Identification of a general transcription factor TFIIAalpha/beta
RT homolog selectively expressed in testis.;
RL J. Biol. Chem. 274:18040-18048(1999).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS, HEART, PLACENTA,
CC KIDNEY, PROSTATE AND UTERUS.
CC -!- SIMILARITY: SOME, IN THE N-TERMINUS TO MEMBERS OF THE ADAPTER
CC COMPLEXES MEDIUM SUBUNITS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF026169; AAD39617.1;
CC HSSP: P32773; 1YTF.
CC MIM: 605357;
CC InterPro: IPR001392; Clathrn_med.
CC InterPro: IPR004855; TFIIA.
CC Pfam: PF00928; Adap_comp_sub; 1.
CC Pfam: PF03153; TFIIA; 1.
CC Transcription regulation; Nuclear protein.
FT DOMAIN 44 150 PRO/SER/THR-RICH.
FT DOMAIN 533 536 POLY-VAL.
SQ SEQUENCE 1182 AA; 132037 MW; 22AD66366D21C371 CRC64;

Query Match
Best Local Similarity 4.9%; Score 106; DB 1; Length 1182;
Matches 72; Conservativity 22.9%; Pred. No.13;

QY 128 QIRHSDPESDFYRGK-GEPTVLSW-HSCRQLLYQAVATILAHAG-----FDCAN 175
Db 750 KVLQSKATEDFFNSIQSLPLTLQLPHSLHQTLQSTASLVIPAGRTLSFTAEIUGTSN 809
QY 176 ES-----VLETLDVAHEYCKLFTKLLRFVAVDREARLGQTFFPDVMEQV 219
Db 810 SSANFTFGYPIHVPAGVLTQVS--GHLYKVNVPIMV--TETSGRAGILQHP--IQOV 862
QY 220 FHEVGIGSVLSLQKFWHRIKIDYHMYLQISQLSEYERYINPEKATEDAKPVKIK--E 277
Db 863 FOOLGQPSVI-----QTSVPQLNPWSLQATTEKSQRIETVLQ-QPAILPSGPFVDRKHLE 915
QY 278 EPVSDITFPVSE-----LEADLASGQD-----SLPMGVLGAQSERPPSNLEVEASPCASSA 329
Db 916 NATSDILVSPGNEHKIVPALICHQESSHYISLPGVFPQVSQTSNDSVESVLSGSASNA 975
QY 330 EYNASPLNLAHVMEPQSEEGNVSGHCVL-----GSDVFEFPMGSM 372
```

```
Db 976 Q-----NL-----HDESLSTSPHGALHQHVTDIQLHILKRMRYGDSVKQPRNIE 1020
QY 373 SEAGIQSPDDSDS 386
Db 1021 EPSNIPVSEKDSNS 1034

RESULT 12
ANK3_HUMAN STANDARD; PRT; 4377 AA.
AC Q12955;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ankyrin 3 (ANK-3) (Ankyrin G).
GN ANK3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain stem;
RX MEDLINE=95138209; PubMed=7836469;
RA Kordeli E., Lambert S., Bennett V.;
RT "Ankyrin. A new ankyrin gene with neural-specific isoforms localized
RT at the axonal initial segment and node of Ranvier.";
RL J. Biol. Chem. 270:2352-2359(1995).
CC -!- FUNCTION: Membrane-cytoskeleton linker.
CC -!- ALTERNATIVE PRODUCTS: A number of isoforms are produced by
CC alternative splicing.
CC -!- TISSUE SPECIFICITY: Expressed in brain and other tissues.
CC -!- SIMILARITY: CONTAINS 23 ANK REPEATS.
CC -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U13616; AAA64834.1;
CC HSSP: P55273; 1BI8.
CC Genew: HGNC:494; ANK3.
CC MIM: 600465;
CC InterPro: IPR002110; ANK.
CC InterPro: IPR000488; Death.
CC InterPro: IPR000906; ZU5.
CC Pfam: PF00023; ank; 24.
CC Pfam: PF00531; death; 1.
CC Pfam: PF00791; ZU5; 1.
CC PRINTS: PR01415; ANKYRIN.
CC SMART: SM00248; ANK; 21.
CC SMART: SM00005; DEATH; 1.
CC SMART: SM00218; ZU5; 1.
CC PROSITE: PS50088; ANK_REPEAT; 21.
CC PROSITE: PS50297; ANK_REPEAT_REGION; 1.
CC PROSITE: PS50017; DEATH_DOMAIN; 1.
CC Cytoskeleton; Alternative splicing; Repeat; ANK repeat.
FT REPEAT 73 102 ANK 1.
FT REPEAT 106 135 ANK 2.
FT REPEAT 139 168 ANK 3.
FT REPEAT 172 201 ANK 4.
FT REPEAT 203 230 ANK 5.
FT REPEAT 234 263 ANK 6.
FT REPEAT 267 296 ANK 7.
FT REPEAT 300 329 ANK 8.
FT REPEAT 333 362 ANK 9.
FT REPEAT 366 395 ANK 10.
FT REPEAT 399 428 ANK 11.
```

```
FT REPEAT 432 461 ANK 12.
FT REPEAT 465 494 ANK 13.
FT REPEAT 498 527 ANK 14.
FT REPEAT 531 560 ANK 15.
FT REPEAT 564 593 ANK 16.
FT REPEAT 597 626 ANK 17.
FT REPEAT 630 659 ANK 18.
FT REPEAT 663 692 ANK 19.
FT REPEAT 696 725 ANK 20.
FT REPEAT 729 758 ANK 21.
FT REPEAT 762 791 ANK 22.
FT REPEAT 795 825 ANK 23.
FT DOMAIN 1519 1898 SER-RICH.
FT DOMAIN 4090 4174 DEATH.
SQ SEQUENCE 4377 AA; F42379E55768B684 CRC64;

Query Match 4.9%; Score 105.5; DB 1; Length 4377;
Best Local Similarity 18.1%; Pred. No. 81;
Matches 91; Conservative 81; Mismatches 169; Indels 161; Gaps 22;

Oy 13 SSSQTNRSSFDLPREFRLVEVDPPLHOPSAKPKPTMLDIPSEP-CSLTITHTIOLIQ 71
Db 2311 TSAQHAEDKQMKFLERILIEVH---TEKGQAEPTEVIRETKKHKEKEMVYVKDLSR 2367
Oy 72 HNRRLRLNI-----ATAQAQNOQQTGKVTKESEPLPSCPGSPPLPDDLPLDCKNPN 124
Db 2368 GDINKDLPEKHDFAFCSEEGQOQEEELTAES--LPSY-----LESSRVN 2413
Oy 125 APF-QIRHSDPES-----DYRKGKGPVTELSNHS-----CRQLLYQAVATILA 167
Db 2414 TPVSOEEDSRSSAQLISDDSYK---TLKLLSQHSIEYHDELSLGRGESYRFAEKMLL 2469
Oy 168 HAGPDCANESVLETLTDVAHEYCLKFTLLRFAYDREAR--LGOTPPDPVMEQVFHEV-- 223
Db 2470 SEKLDVSHSDTEESTVDHAGPPSELO-----GSDKRSREKIAATAPKEILSKYKDYSE 2524
Oy 224 -GIGSVLSLQKFWHRIKDY-----HSTMQLQSK-QLSEEVERIVNPKATEDAKPV 273
Db 2525 NGGVKVSDEHFDKVTVLVHSGNVSPPKHAMMRFTEDRLDRGREKLIYEDRVDTVKEA 2584
Oy 274 KIKEEPVS---DITFPVSELEAD-----LASGD 299
Db 2585 EEKLTVEVSQFTRDKTEKLNDELQPEKKARPKNKEYSSQSPTSSSPKVLTTLLASND 2644
Oy 300 Q-----SLPMG-----VLGAQSERFPNLEAVEASQASSA----- 329
Db 2645 EMVKARQHPDQGFPAERKAPSLPSSPEKMWLSQQTEDSKSTVEAKGSISQSKAPDGP 2704
Oy 330 ----EVNASPLWNL-----AHVKMEPQSEEGNVSGHGVLSGVDFEPPMSGHSEAGIP 378
Db 2705 QSGFLQKQSLSSIRLKFQETHAKSKDMSQEDRKSDQSRI-----PVKKIQESKLP 2757
Oy 379 -----QSPDSDPS 386
Db 2758 VYQVFAREKQKKAIDLDPDESVS 2779
```

RESULT 13

```
PAC2_RAT STANDARD; PRT; 488 AA.
AC Q9QY17; Q9QY18; Q9QY19; Q9QY20;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein kinase C and casein kinase substrate in neurons 2 protein
DE (Synaptic dynamin-associated protein II) (Syndapin 2) (Syndapin-II)
DE (SdpII).
GN PACSIN2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_taxid-10116;
RN [1]
```

```
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).
RC STRAIN-Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=20171438; PubMed=10704453;
RA Qualmann B., Kelly R.B.;
RT "Syndapin isoforms participate in receptor-mediated endocytosis and
actin organization.";
RL J. Cell Biol. 148:1047-1062(2000).
RN [2]
RP FUNCTION.
RX PubMed=11082044;
RA Modregger J., Ritter B., Witter B., Paulsson M., Plomann M.;
RT "All three PACSIN isoforms bind to endocytic proteins and inhibit
endocytosis.";
RL J. Cell Sci. 113:4511-4521(2000).
CC -!- FUNCTION: May play a role in vesicle formation and transport.
CC -!- SUBUNIT: Homo- and hetero-aggregates with other PACSINS. Binds
dynamitin 1, synaptotagmin, synapsin 1 and the neural Wiskott-Aldrich
syndrome protein (N-WASP).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Vesicle-like cytoplasmic
distribution.
CC -!- ALTERNATIVE PRODUCTS: 4 isoforms; 1/aa (shown here), 2/ab, 3/ba
and 4/bb; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed. Isoforms 1 and 3 are
expressed in PC12 cell line and heart, whereas isoforms 2 and 4
are seen in most tissues examined with higher levels of expression
in muscle, testis and brain.
CC -!- PTM: Phosphorylated by casein kinase 2 (CK2) and protein kinase C
(PKC) (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PACSIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 FCH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; AF139492; AAF22211.1; -
DR EMBL; AF139493; AAF22212.1; -
DR EMBL; AF139494; AAF22213.1; -
DR EMBL; AF139495; AAF22214.1; -
DR HSSP; P29355; 1SEM.
DR InterPro; IPR001060; Cdc15_Fes_C1P4.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00611; FCH; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00055; FCH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS01033; FCH; 1.
DR PROSITE; PS00002; SH3; 1.
DR Endocytosis; SH3 domain; Coiled coil; Phosphorylation;
KW Alternative splicing.
FT DOMAIN 11 75 FCH.
FT DOMAIN 428 488 SH3.
FT DOMAIN 184 239 COILED COIL (POTENTIAL).
FT VARSPPLIC 302 303 MISSING (IN ISOFORM 3 AND ISOFORM 4).
FT VARSPPLIC 346 386 MISSING (IN ISOFORM 2 AND ISOFORM 4).
SQ SEQUENCE 488 AA; 55978 MW; B2975012EF0DF56 CRC64;

Query Match 4.8%; Score 105; DB 1; Length 488;
Best Local Similarity 22.5%; Pred. No. 4.5;
Matches 71; Conservative 33; Mismatches 103; Indels 108; Gaps 13;

Oy 167 AHAGFDCANESVLETLTDV-----AHEYCLKFTLLRFAYDREA----- 205
Db 134 AEDGFRKQKFWAKKLEVDAAKKAHTACKKEL--AVSREANSKADPSINPQLAKL 190
Oy 206 -----RLGQT--PFDPVMEQVFHEVGIGSVLSLQKFWHRIK 240
```

```
Db 191 ODKIEKQDVLTKDKYKALKKELDQTTPOYMENMEQVFEQ-----COOFEEKRLR 242
Qy 241 DYHSYMLQISKOLS-----EYERIVNPEKATEDAKPVKIKEEVPSDITTP--V 287
Db 243 FFREVLEVOXKHLDSNVASYKGIYRELESGIKAADAVEDLRFWRANHGFCGMANNWPFQ 302
Qy 288 SEELEADL-----ASGDQSLPMGVGLGAQS--ERFPNSLEVEASPOASSARVNA 333
Db 303 DEWSADLNLRTLSRREKKKAADGVTLTGINOTGDSQGNKPSNLSVPSNP-AQSTOLQS 361
Qy 334 SPLWNLAHVKMEQSEEGNVSGHVLGSDVFEPP-----MSGMSEAGIPQSDSDS 386
Db 362 S-----YNPFEDDDT-----GSSVSEKEDIKAKNVSSYKTONYPADWSDDES 405
Qy 387 SYGSHSTDSLWGSSP 401
Db 406 NPFSTSDANGDSNP 420

RESULT 14
Y249_HUMAN STANDARD; PRT; 896 AA.
AC Q2539;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 35, Last sequence update)
DE Hypothetical protein KIAA0249.
GN KIAA0249
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=97191544; PubMed=90399502;
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
CC -1- SIMILARITY: TO HUMAN KIAA0188 AND YEAST SMP2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D87436; BAA13380.1; -
DR Genbank; HGNC:14450; LPIN2.
KW Hypothetical protein.
SQ SEQUENCE 896 AA; 99399 MW; 080113FCCA533272 CRC64;

Query Match 4.8%; Score 105; DB 1; Length 896;
Best Local Similarity 19.1%; Pred. No. 10;
Matches 76; Conservative 56; Mismatches 147; Indels 118; Gaps 16;

Qy 7 WGEIPITSSQTNRSSFDLLPR-----EPRVVEHDPPLHOPSAN----- 45
Db 266 WGGFPSTKVKRSRSDHPTATITPSENTHRVIPSDNLISEVEKASMEDTIVCTIV 325
Qy 46 KKPPTMLDIPSPCSLTHTIQLIQ---HNRLRLNLIATAQNOQOQTEGVKTERSEPL 102
Db 326 KKPRLALGTQMSDPTSA---ELLEPPLESTQISSMLDADHLPNLALEA---PSESK 377
Qy 103 PSCPGSPPL-----PDDLPLDCK--NPNAPE---FQIRHSD-----PESDF 138
Db 378 PAAKVDSPSKKGVHKKRSQHQGPDIDYLDLGLPEVAALYFPKSESPGSRQWPESDT 437
```

```
Qy 139 YRGKGPVTELSWHSCRLQLYQAVATILAHAGFCANESVLETTDVAHEYC----- 190
Db 438 LSGSQSP-----QVGSAAADSGTECLSDSAMD-LPDVTLSLGGLSNGE 482
Qy 191 ---LKTFTKLRFADVAREARGQTFPPDMVQVPEHE-----VGIGSVLSLQKFWOHRIDY 242
Db 483 ISKEKFEHITITTHFAENPGLIDNPNLVIRIYNRYNVALAAPMILSLQVFK----- 536
Qy 243 HSYMLQISKOLSEYERIVNPEKA-----TEDAKPVKIKEEVPSDITFPV 287
Db 537 -----SLPKATVSEWVKDKMPKKSGRWFWFKRKRESMTKQLPESKEGKSEAPPASDLPPSS 591
Qy 288 SEELEADLQSGDQSLPMGVGLGAQSRRFPNSLEVEASP 324
Db 592 KEPAGARPAENDSSDEG-----SQELEESITVDPIP 623

RESULT 15
SALL_HUMAN STANDARD; PRT; 1324 AA.
AC Q9NSC2; Q9NSC3; Q99881; Q9P1R0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Sal-like protein 1 (Zinc finger protein SALL1) (Spalt-like
DE transcription factor 1) (HSall).
GN SALL1 OR SALL1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., DISEASE, AND VARIANTS SER-150 DEL; SER-150 INS AND
RP GLY-159.
RX MEDLINE=99138656; PubMed=9973281;
RA Kohlhasse J., Taschner P.E.M., Burfeind P., Pasche B., Newman B.,
RA Blantic C., Breuning M.H., ten Kate L.P., Maaswinkel-Mooy P.,
RA Mitulla B., Seidel J., Kirkpatrick S.J., Pauli R.M., Wargowski D.S.,
RA Devriendt K., Proesmans W., Gabrielli O., Coppa G.V.,
RA Wesby-van Swaay E., Trembath R.C., Schinzel A.A., Reardon W.,
RA Seemanova E., Engel W.;
RT "Molecular analysis of SALL1 mutations in Townes-Brocks Syndrome.";
RL Am. J. Hum. Genet. 64:435-445(1999).
RN [2]
RP SEQUENCE OF 1-26 FROM N.A., DISEASE, AND VARIANTS SER-164 DEL AND
RP GLU-1265.
RX MEDLINE=20004537; PubMed=10533063;
RA Marlin S., Blanchard S., Lacombe D., Denoyelle F., Alessandri J.-L.,
RA Calzolari E., Drouin-Garraud V., Ferraz F.G., Fourmaintraux A.,
RA Philip N., Toubanc J.E., Petit C.;
RT "Townes-Brocks syndrome: detection of a SALL1 mutation hot spot and
RT evidence for a position effect in one patient.";
RL Hum. Mutat. 14:377-386(1999).
RN [3]
RP SEQUENCE OF 26-1324 FROM N.A.
RX MEDLINE=97131507; PubMed=8975705;
RA Kohlhasse J., Schuh R., Dove G., Kuehnlein R.P., Jaeckle H.,
RA Schroeder B., Schulz-Schaeffer W., Kretzschmar H.A., Koehler A.,
RA Mueller U., Raab-Vetter M., Burkhardt E., Engel W., Stick R.;
RT "Isolation, characterization, and organ-specific expression of two
RT novel human zinc finger genes related to the Drosophila gene spalt.";
RL Genomics 38:291-298(1996).
RN [4]
RP SEQUENCE OF 313-345 FROM N.A., AND DISEASE.
RX MEDLINE=98085876; PubMed=9425907;
RA Kohlhasse J., Wischermann A., Reichenbach H., Froster U., Engel W.;
RT "Mutations in the SALL1 putative transcription factor gene cause
RT Townes-Brocks syndrome.";
RL Nat. Genet. 18:82-83(1998).
RN [5]
RP DISEASE.
RX MEDLINE=20381976; PubMed=10928856;
```

RA Engels S., Kohlase J., McLaughran J.;
RT "A SALL1 mutation causes a branchio-oto-renal syndrome-like
RL J. Med. Genet. 37:458-460(2000).
CC -!- FUNCTION: Transcriptional repressor involved in organogenesis (By
CC similarity).
CC -!- SUBUNIT: Interacts with HDAC1, HDAC2, RBBP4, RBP7, MTA1 and MTA2
CC complex (HDAC).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- TISSUE SPECIFICITY: Highest levels in kidney. Lower levels in
CC adult brain (enriched in corpus callosum, lower expression in
CC substantia nigra) and liver.
CC -!- DEVELOPMENTAL STAGE: In fetal brain exclusively in neurons of the
CC subependymal region of hypothalamus lateral to the third
CC ventricle.
CC -!- DISEASE: Defects in SALL1 are the cause of Townes-Brocks syndrome
CC (TBS); a rare, autosomal dominant malformation syndrome with a
CC combination of imperforate anus, triphalangeal and supernumerary
CC thumbs, malformed ears and sensorineural hearing loss.
CC -!- DISEASE: Defects in SALL1 are in rare cases associated with
CC hemifacial microsomia (HFM) or Goldenhar syndrome; a common birth
CC defect involving first and second branchial arch derivatives. The
CC highly variable phenotype shows in addition to craniofacial
CC anomalies cardiac, vertebral, and central nervous system defects.
CC -!- DISEASE: Defects in SALL1 are associated with branchio-oto-renal
CC (BOR) syndrome, an autosomal dominant disorder manifested by
CC various combinations of preauricular pits, branchial fistulae or
CC cysts, lacrimal duct stenosis, hearing loss, structural defects of
CC the outer, middle, or inner ear, and renal dysplasia. Associated
CC defects include asthenic habitus, long narrow facies, constricted
CC palate, deep overbite, and myopia. Hearing loss may be due to
CC Mondini type cochlear defect and stapes fixation.
CC -!- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y18265; CAB41400.1; -.
CC DR EMBL; Y18264; CAB41399.1; -.
CC DR EMBL; X98833; CAB41399.1; JOINED.
CC DR EMBL; AF017655; AAB99908.1; -.
CC DR EMBL; AF074949; AAF19263.1; -.
CC DR HSP; P15822; IBB0.
CC DR Genew; HGNC:10524; SALL1.
CC DR MIM; 602218; -.
CC DR MIM; 107480; -.
CC DR MIM; 164210; -.
CC DR MIM; 113650; -.
CC DR InterPro; IPR000822; znf_C2H2.
CC DR Pfam; PF00096; zf-C2H2; 10.
CC DR Prodom; PD000003; znf-C2H2; 1.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
CC DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 9.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW Zinc-finger; Metal-binding; Repeat; Polymorphism.
FT ZN_FING 449 471
FT ZN_FING 477 499
FT ZN_FING 477 499 C2H2-TYPE.
FT ZN_FING 706 728 C2H2-TYPE.
FT ZN_FING 734 756 C2H2-TYPE.
FT ZN_FING 766 788 C2H2-TYPE.
FT ZN_FING 1001 1023 C2H2-TYPE.
FT ZN_FING 1029 1051 C2H2-TYPE.
FT ZN_FING 1134 1156 C2H2-TYPE.
FT ZN_FING 1162 1184 C2H2-TYPE.
FT ZN_FING 150 159 POLY-SER.
FT DOMAIN 160 163 POLY-GLY.

FT	DOMAIN	237	240	POLY-GLN.
FT	DOMAIN	294	297	POLY-ALA.
FT	DOMAIN	371	375	POLY-SER.
FT	DOMAIN	1144	1147	POLY-SER.
FT	VARIANT	150	150	MISSING.
FT	VARIANT	150	150	/FTid=VAR_013155.
FT	VARIANT	150	150	S -> SS.
FT	VARIANT	159	159	/FTid=VAR_013156.
FT	VARIANT	159	159	S -> G.
FT	VARIANT	164	164	/FTid=VAR_013157.
FT	VARIANT	164	164	MISSING.
FT	VARIANT	1265	1265	/FTid=VAR_013158.
FT	VARIANT	1265	1265	G -> E.
FT	VARIANT	79	79	/FTid=VAR_013159.
FT	CONFLICT	79	79	G -> A (IN REF. 1; CAB41399).
SQ	SEQUENCE	1324	AA; 140390	MM; CBE787847A96D8AA CRC64;

Query Match 4.8%; Score 103.5; DB 1; Length 1324;
Best Local Similarity 22.0%; Pred. No. 22;
Matches 87; Conservative 53; Mismatches 140; Indels 115; Gaps 20;

QY	34	VHDPPLHQP	SANKPKPTMLDIP	PCSLTTHTI	-----OLIQHN--RRLRNLIATAQA	85	
DB	74	VNENPGSP	PETFSPPP	--DNP	EQMNDTVNKTDQVDCSLSEHGLDRESEVEAPV	131	
QY	86	QNOQTEGV	KTESEPLP	SCPGSP	PLPDDLPLDCKNPNAPFQIRHSDPESDFYRGKEP	145	
DB	132	ANK	---SGSGT	SSGSHS	STAPSS-----SSSSSSSSSGGGSS	165	
QY	146	VTLSWHSC	ROLLYQAVATIL	AHAG--FDC	ANESVLETLTDVAHEYCLKFKTLRFAVDR	203	
DB	166	STGTS	-----AITS	LPQLGLTTLGNF	SVINSNVIEN---LQSTKVAQFSQ	212	
QY	204	EARL	-----GOT	PPDVMEQVFEV	GIGSVLSLQKFWQHRKQYHSMQLQSKOLSEYE	258	
DB	213	EARGGAS	GKLA	VPALMEQ	-----LLALQ---QQQI-----HQLQIEIRHQIL	255	
QY	259	RIVN	-----PEK	ATEDAKPVK	KEEPVSDITFPVSEELADLASC-DOSLPMGVLAGQSE	312	
DB	256	LLASON	ADLP	TSSSPSQ	GTLRTSANPLSTLSSHLSQLAA--AAGLAQSL	303	
QY	313	RPNSNLE	VASPOASSA	EVNASPLW	NLAHYKMEQSEEGNMGVGLSDVFPEPMSCM	372	
DB	304	-----AS	QASISIGV	KQLPPIQL	-----PQSS-----SGNTIPNSGSSPNMNI	343	
QY	373	SEAGI	---POS	DDSDSYGSH	STDLSM--GSSPVF	403	
DB	344	LAAAVT	PSSEK	VASSAGAS	HVSNPVS	SSSSSPAF	378

Search completed: July 18, 2003, 09:57:50

Job time : 31.5071 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	Match				
1	142.5	6.6	6.6	979	2	T40066	hypothetical prote
2	126	5.8	5.8	1621	2	A82255	hypothetical prote
3	123	5.7	5.7	552	1	A31401	macrophage colony-
4	123	5.7	5.7	552	1	S33703	colony-stimulating
5	121.5	5.6	2957	2	S37302	hypothetical prote	
6	117	5.4	575	2	A69716	spore coat assembl	
7	112.5	5.2	1332	2	A41552	probable transcrip	
8	112	5.1	609	2	T28736	hypothetical prote	
9	110.5	5.1	785	2	S54016	SOK2 protein - yea	
10	108.5	5.0	671	2	S53407	CHS5 protein - yea	
11	108.5	5.0	1206	2	T34021	protein kinase SK2	
12	107	4.9	2484	2	T26216	hypothetical prote	
13	107	4.9	2607	2	T26215	hypothetical prote	
14	106	4.9	387	1	A43704	involucrin - wester	
15	106	4.9	1180	2	T20773	hypothetical prote	
16	106	4.9	4549	2	T20771	hypothetical prote	
17	106	4.9	4667	2	T20774	hypothetical prote	
18	106	4.9	5170	2	T15348	hypothetical prote	
19	105.5	4.9	444	2	T25111	hypothetical prote	
20	105.5	4.9	4377	2	A55575	ankyrin 3, long sp	
21	105	4.8	1163	2	A36685	205K microtubule-a	
22	104.5	4.8	727	2	A89131	protein F10G7.9 [l	
23	104.5	4.8	834	2	D57282	ankyrin-related pr	
24	104	4.8	1186	2	S72229	meiotic recombinat	
25	103.5	4.8	914	2	T21122	hypothetical prote	
26	103.5	4.8	2224	1	KFHU5	coagulation factor	
27	103	4.7	1403	1	A47328	natural killer cell	
28	103	4.7	1826	2	H86502	exonuclease ABC s	
29	102.5	4.7	679	2	T24365	hypothetical prote	

Query Match 5.6%; Score 121.5; DB 2; Length 2957;
Best Local Similarity 20.0%; Pred. No. 4.5;
Matches 86; Conservative 44; Mismatches 172; Indels 127; Gaps 17;

QY 37 PPLHQPANKPK-----PPTMLDIPSEPCSLTHTIOLIOHNRRLRLNIATAQ 84
DB 132 PPPHPYSSNGSQGYRGPVNTGPPPGQYPPPP--QAPQNYQOMYHOOR-----AQQQ 184
QY 85 AQNOQQTEGVKTESEPLPCP-----GSPPLPDLLPLDCKKNPNAPQIRHSDPES----- 136
DB 185 QAGQOQAAG--PPCYGYGYAPQASTGYPPPPSOQSPY-APQAGPPQMRHQYPPHQQQ 241
QY 137 -----DFYRKGEFVTWELSHWSQRLLYQAVATILAHAGFDCANESVLETLTDVAHEYC 190
DB 242 APPGYWDGYQGYGPPPS-----QQQQQGGGVPVTAQSQMQMAQQ----- 281
QY 191 LKFTKLLRFAVDREARLQTFPPDVMQVPH-----EVGIGSVLSLQKF-----WQ 236
DB 282 -----EQMGRVGTNDLMNVIAQVFKSSAECKARIQQLYNNMORSASADAEOQLQ 330
QY 237 HRIKDYHSYMLQISKQLSE-EYERIVNPEKATEDAKPVKIEEPVSDITFPVSEELADL 295
DB 331 HKLAQFOSDHYRYSQQQHQWQQQQQQQAAAA-----AAAAAAS 373
QY 296 ASGDQSLLPMGLGAQSERFSPN-----LEVEASPOASSAEVNASPLWNLAHKMEPQSEEG 352
DB 374 AGGHQGGPPPTSSQSQWIFSNPNVPVHVNYPCGSQVHIITADKTRTLISVYHD----- 426
QY 353 NVSGHGVLGSD-----VFPEPMSGSEAGIPOSDD-----SDSSYSGSHSTOSLM 397
DB 427 ---GYGAGAGSASTESKESIVPPEPTPSMQESVPVNAQVKMEPEKPSPYQQQYNGMEN 483
QY 398 GSPVPVFNOR 406
DB 484 GSGNPYAAQ 492

RESULT 6
A69716
Spore coat assembly protein (spovID) - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: A69716; B47083
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emmerson, P.F.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinc
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, C.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portele
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scan
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstora, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: A69716
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-575 <KUN>
A:Cross-references: GB:Z99118; GB:AL009126; NID:g2635200; PIDN:CAB14771.1; PID:ell840
A:Experimental source: strain 168
R:Beall, B.; Driks, A.; Losick, R.; Moran Jr., C.P.
J. Bacteriol. 175, 1705-1716, 1993
A:Title: Cloning and characterization of a gene required for assembly of the Bacillus
A:Reference number: A47083; MUID:93194796; PMID:8449878
A:Accession: B47083
A:Status: preliminary
A:Molecule type: nucleic acid

Db 552 ASPRH-----QEE-----DIELEAPKDATSVAVEP--SNEDVKPEEKGEAEDDI 596
Qy 352 GNVSGHGVLDVPEEPMGMEAGIPQSPDDSDSGSHSTDSLM-----GSSPV 402
Db 597 NNVSKAASGESTHOKTEASASLESSAVTEQEETEAENVTDVLTSTKEAKKNTGNSN 656
Qy 403 FNORCKKMRK 413
Db 657 NKKNKKKKKK 667

RESULT 11
T34021
protein kinase SK2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34021
R:Fukami, Y.; Yamamoto, H.; Ichihara, T.; Mori, K.; Gomi, T.; Sato, K.
submitted to the EMBL Data Library, April 1997
A:Description: SK2, a putative rat homologue of yeast protein kinase NRK1.
A:Reference number: 221463
A:Accession: T34021
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1206 <FUK>
A:Cross-references: EMBL:AB003357; NID:d1106513; PID:d1020890; PIDN:BA20077.1
A:Experimental source: strain SD
C:Genetics:
A:Gene: SK2

Query Match 5.0%; Score 108.5; DB 2; Length 1206;
Best Local Similarity 20.3%; Pred. No. 11;
Matches 128; Conservative 68; Mismatches 180; Indels 255; Gaps 31;

Qy 6 YWGEIPSSSTNRS-----SFDLLPREFRIVEVH--DPLLHQPS-----ANKPKPPT 51
Db 195 YWMAPEVVMCTSKDRPYDKADVMSGLITLIEMAEIPEPHHNLPMRWLLKIAKSEPT 254
Qy 52 MLDIPSPCSLTIH-----TIQLOH-----NRRRLNLATAQAQ 86
Db 255 L-----AQSRWSSNFKDFLKCLEKNVDARWTTSQLQHPVTVDSNKNPVRELLAEAK 310
Qy 87 NQOQTEGVK-----TESEPLP-----NRRRLNLATAQAQ 103
Db 311 VTEEDGKEDEDDDETESALPIANKRASSDLSIASEEDKLSQACILLESVSERTEHN 370
Qy 104 -----SCPGSPPLPDLLPD--CKNPNAPFOIRHSDPESDFY---RGKGPEVT 147
Db 371 TSGDKFSNKLVSERKTPGEPKTVDDGPDANDVNLETVAEPNDQAVGFHENGREKRPQL 430
Qy 148 ELSWHSCROLLYQAVATILAHAGFDCANESVLETLTDAHEYCLKFTKLLRFAYDR--- 203
Db 431 ESQPTEDQ---QTVDNLVGEGND--SNVILETNTD-----CLKPEE-----DRNEEN 475
Qy 204 ----EARLQGT--PPFDV-----MEQVFHEVGIGSVLSLQKFWQHRK 240
Db 476 QEILNKLTSQEEIKDHIQTMDLVSQETGEKADFQAINDEVGFTKEETQEKLG----K 531
Qy 241 DYHMYMLQISKOLSE-----EYERINVP-----EKATEDAKP 272
Db 532 DDKTHKVVISDITSEVGTDEPPGDTOKSAEQSDAEGGAGEAPEAQTUTEKATEGPEA 591
Qy 273 VKIKEEP-----VSDITFPVSEEL-----EADLASGDQSLP 303
Db 592 HGAEEEPGRGRVEDKQEQSAVCEGQVTSSTESTRATTEPTEPTEVDQVQSESNIE 651
Qy 304 ----MGVLGAQSERFSS-----NLEVASPQASAEVNASPLNLAHVKMEPO----- 347
Db 652 ELERLGTGAEEQALGSKGAATLDERENAQELPVKAEPQAPAAQSEPPVPLIPS 711
Qy 348 ---ESEGNGVSGHGVLDVPEEPMGMEAGIPQSPD-----DSDSSYGS---HSTDLS 936
Db 936

Db 712 INIHSE--NTENKKGEMGA--LPKP-----ETILPEPEPENGKNGNDTSGTGSTVENSSDL 762
Qy 397 MGSSPVF-----NORCKKMRK 413
Db 763 NLSISSFLSKTKDSGSVSLQETRRQKTLKK 793

RESULT 12
T26216
hypothetical protein W06A7.3c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26216
R:Ainscough, R.
submitted to the EMBL Data Library, August 1996
A:Reference number: 220173
A:Accession: T26216
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2484 <WIL>
A:Cross-references: EMBL:Z78066; PIDN:CAB51467.1; GSPDB:GNO0023; CESP:W06A7.3c
A:Experimental source: clone W06A7
C:Genetics:
A:Gene: CESP:W06A7.3c
A:Map position: 5
A:Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2339/2; 2463/2

Query Match 4.9%; Score 107; DB 2; Length 2484;
Best Local Similarity 19.2%; Pred. No. 42;
Matches 85; Conservative 70; Mismatches 168; Indels 120; Gaps 20;

Qy 28 EFRVLEVDHPLHOPSAANKPKPTMLDIPSPCSLTITHTIQLIQHNRRRLNLATAQAQ 87
Db 1317 EQSIANWIDELVHEDDEKK-----VPEVTANISVSASENID--DSTTANAVKTEVSS 1367
Qy 88 QQ-----QTEGVKTEES-----EPL-----PSCPGSPPLPDLLPL----- 118
Db 1368 EQLQVATVFELESAPAEESAAIPEVQEPLEKVEVQPDLSQNSAPAHKIIDLHFNIPKDH 1427
Qy 119 -DCKNPNAPFOIRHSDPESDFYCKGE-----PVTLSWHSCRQ-----LLYQAVATI 165
Db 1428 EDYGNDRVPGTESSE--ESQKADGNQENQEEEDVVAELNFIHQWDEDEVISLQSLKSL 1486
Qy 166 LAHAGFDCANESVLETLTDAHEYCLKFTK-----LLRFADVREARLGTPTFPDVMQ 218
Db 1487 VAEVG--CITDVSADVNEQDEESTLKLKWPSEPSLELDFTNPKVIHVPIP-LMEP 1543
Qy 219 VFHEVGIGSVLSLQKFWOHRKIDYHMYMLQISKOLSEYERINVPKATEDAKPVKTK-- 276
Db 1544 -----ATMYLEEMVEWIIAD-----AVKEVSE--MEVVTSEISEMAPQVSESTC 1586
Qy 277 --EPPVSDITFPVSEELAADLASGDQSLPMGLGAQSERFSPNLEVASP----- 324
Db 1587 PIPEPLADLKLPLVDEDEKT-----PEPEVPVPGVQOERIIPIEVQAPQITPQRPAP 1639
Qy 325 -----QASSAEVNASPLNLAHVKMEPOEESGVSHGVLSGSDVFEPEPMGMS 373
Db 1640 KSELPKVAKPLDDSKSRVREAPL-----NKLGRYSEEQKE-----LVESLERPLTIT 1690
Qy 374 EAGIPQSPDSDSGSHSTDSL 396
Db 1691 QQKPEKPTEDIGALSPLSPNTL 1713

RESULT 13
T26215
hypothetical protein W06A7.3a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26215
R:Ainscough, R.
submitted to the EMBL Data Library, August 1996
A:Reference number: 220173

A:Accession: T26215
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2607 <WIL>
A:Cross-references: EMBL:Z78066; PIDN:CAB01522.2; GSPDB:GN00023; CESP:W06A7.3a
A:Experimental source: clone W06A7
C:Genetics:
A:Gene: CESP:W06A7.3a
A:Map position: 5
A:Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2412/1; 2462/2; 2586/2

Query Match 4.9%; Score 107; DB 2; Length 2607;
Best Local Similarity 19.2%; Pred. No. 45;
Matches 85; Conservative 70; Mismatches 168; Indels 120; Gaps 20;

QY 28 EFRLEVDHDPHOPSSANKPKPTMLDIPSECSLTHTTQLQHNRLRLNLIATAQON 87
DB 1317 EQSIANVIDELVHEDDEKK-----VPEVTANISVASENDID-DSTTANAVPKTEVSS 1367

QY 88 QQ-----OTEGVKTEES-----EPL-----PSCGSPPLPDDLLPL----- 118
DB 1368 EQLQVATVEFELESAPESAAIPEVQPLEKVEVQPDLSNAPAPHKIIDLHNIPKDH 1427

QY 119 -DCKNPAPQIRHSDPESDFYCKGE-----PVTLSWHSCRQ-----LLYQAVATI 165
DB 1428 EDYGNVYVPFGTSESE-ESQKADGNOENQEBEDVVAELNFPPIQWRDEDVISLSKSL 1486

QY 166 LAHAGFCANESVLETLTDVAHEVCLKFTK-----LLRFVAVDREARLGOTFPFDVMEQ 218
DB 1487 VAEVG--CITVDVADSDNEQDESTLKILKVPSEPSLLELDFTNPKVIHVPIP-LMEP 1543

QY 219 VFHEVGISVLSLQKFWOHRIDYHSYMLQISQKSEYERIVNPEKATEDAKPVKIK-- 276
DB 1544 -----ATWYLEEMVEWIIAD-----AVKEVSE--MEVTESEISEMAPOVSESTC 1586

QY 277 --BEPVSDIIFPVEEADLASQDQLPMGVLGAQSERFPNSLEVAESP----- 324
DB 1587 PIPEPLADLPLVDEDEKT-----PEPEPVVPGVQVQRIPIEVEQAFTIPQRPPRAP 1639

QY 325 -----QASSAVBNASPLNLAHVAKMEQSEGNVSGHVLGSDVFPEPMGMS 373
DB 1640 KSELPAKVLDDSKSVRAPL-----NIKLRTYSEEQKE-----LVESLERPLTIIT 1690

QY 374 EAGIQSPDSDSYSGSHSTDLSL 396
DB 1691 QOKPPEKPTEDIGALSPLSPNTL 1713

RESULT 14
A43704
involucrin - western tarsier
C:Species: Tarsius bancanus (western tarsier)
C:Date: 20-Feb-1993 #sequence_revision 23-Feb-1996 #text_change 22-Jun-1999
C:Accession: A43704
R:Djian, P.; Green, H.
Proc. Natl. Acad. Sci. U.S.A. 88, 5321-5325, 1991
A:Title: Involucrin gene of tarsoids and other primates: alternatives in evolution of the linked envelope under the plasma membrane.
A:Reference number: A43704; MUID:91271381; PMID:1905021
A:Accession: A43704
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-387 <DJ>
A:Cross-references: GB:M65124; NID:g343451; PIDN:AAA36960.1; PID:g343452
C:Comment: During the terminal differentiation of keratinocytes, this protein from the linked envelope under the plasma membrane.
C:Superfamily: involucrin
C:Keywords: cornified cell envelope; duplication; epidermis; tandem repeat

Query Match 4.9%; Score 106; DB 1; Length 387;
Best Local Similarity 23.0%; Pred. No. 3.1;
Matches 76; Conservative 37; Mismatches 140; Indels 78; Gaps 17;

QY 77 RNLATAQAOQOQTESEPLFCGSPPLPDDLLPLDCKN-PN---APFQIRHS 132

Db 17 QELLKTVPPANTQQDOMKQTPSPAPCQKPSGSELVPEKHAPVPKQVECEPQOQDHQ 76
QY 133 DPESDFYRGKGEVPELTSWHSCRQLLYQAVATILAHAGFCANESVLETLTDVAHEVCLK 192
Db 77 EPELQGRKQOEP-QEQEVHPGKQ--QOKPQEQEHAHLG-----KKQEPQGEVH----- 122
QY 193 FTKLLRFVAVDREARLGOTFPFDVMEQVFHEVGISVLSLQKFWOHRIDYHSYMLQISQK 252
Db 123 LGRQOQKQTQGEVHLGKQKQ-QELQEQEVH-----LEKQLEP-QEVH-----LEKQ 166
QY 253 LSEYERIVNPEKATEDAKPV-----KIKEEPVSDIIFPVEEADLASQDQLPMGV-- 306
Db 167 LQEQE---VHLEKQLOQEPPELNLGKQOQEP-----QEQEAVLGRKQOQELPEPQDP 214
QY 307 --LGAOSERFPNSLEVAESPOASSARVNASPLNLAHVAKMEQSEGNVSGHVLGSD 363
Db 215 ELHLGKQOQE-PQEQEVQLEKQEQEQE-----LHLGKQOQESQ----- 253
QY 364 VFPEPMGMEAGIPIQSPDSDSYSGSHSTD 394
Db 254 --EQLHLRLKLOQVPOEPQDELHLGKQOQE 282

RESULT 15
T20773
hypothetical protein ZK270.2c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20773; T27816
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19322
A:Accession: T20773
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1180 <WIL>
A:Cross-references: EMBL:Z81499; PIDN:CAB04092.1; GSPDB:GN00019; CESP:ZK270.2c
A:Experimental source: clone F11C3
R:Kershaw, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z20423
A:Accession: T27816
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1180 <WIL2>
A:Cross-references: EMBL:Z82089; PIDN:CAB05005.1; GSPDB:GN00019; CESP:ZK270.2c
A:Experimental source: clone ZK270
C:Genetics:
A:Gene: CESP:ZK270.2c
A:Map position: 1
A:Introns: 933/3; 974/3; 1008/2; 1100/3; 1151/3

Query Match 4.9%; Score 106; DB 2; Length 1180;
Best Local Similarity 19.9%; Pred. No. 16;
Matches 94; Conservative 51; Mismatches 177; Indels 150; Gaps 19;

QY 23 DLLPREFLVEVHDPLHQPS-----ANK-----PKP---PTMLDIPSECSLTIHT 66
Db 182 DIRP-----VVLEAQDPLSDHSQVYHNGESWKAENKIKPPKPVRTKKGUSPEIESEVHE 237

QY 67 IQLIQHNRLRLNLIATAQAOQOQTEGVKTESEPLFCSP-----GSPPLPDDLL 116
Db 238 IRL-----TRVASSSEPEVDVIPAESTVPKFRFLRFLSIGKKTSPDSEF 286

QY 117 PLCKNPAPQIRHSDPESDF-----YRGKEPVTELSWHSCRQLLYQAVATIL 166
Db 287 PECSQVGDGVDLTSRSDLEVMPLYLHGSVPPSPKFTPVKV----- 329

QY 167 AHAGFCANESVLETLTDVAHEVCLKFTKLLRFVAVDREARLGOTP-----FPDVMQVPH 221
Db 330 -----DKADEERKRYTVIYRHEGEEDFADKPAYGFSDDVYTGSLN 372

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 07:48:47 ; Search time 91.237 Seconds
(without alignments)
604.643 Million cell updates/sec

Title: US-09-857-308-1
Perfect score: 2175
Sequence: 1 MNLQRYGWEIPISSTQTRNS.....SLMGSPVFNQCKKMRKI 414

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2175	100.0	414	21	Human ART-1 protei
2	2175	100.0	419	22	Human protein sequ
3	2155	99.1	412	22	Human cell death p
4	2121	97.5	403	21	Gene 41 human secr
5	2059.5	94.7	404	21	Human secreted pro
6	1898	87.3	363	22	Human cell death p
7	1022	47.0	199	22	Human cell death p
8	915	42.1	198	20	Human endometrium
9	858	39.4	169	22	Human cell death p
10	572	26.3	111	22	Human cell death p

11	371	17.1	71	22	AAG98658	Human cell death p
12	239	11.0	46	22	ACG98662	Human cell death p
13	224	10.3	43	22	ACG98663	Human cell death p
14	132	6.1	359	22	ABB62430	Drosophila melanog
15	131	6.0	1049	22	ABB58717	Drosophila melanog
16	123	5.7	520	22	AAE00578	Murine colony stim
17	123	5.7	520	22	AAE00579	Murine colony stim
18	123	5.7	520	22	AAE00580	Murine colony stim
19	123	5.7	552	9	AAE03390	Deduced sequence f
20	123	5.7	552	17	AAW10070	Murine long form C
21	123	5.7	552	18	AAW36145	Mouse long form of
22	123	5.7	552	18	AAW35752	Murine colony stim
23	123	5.7	552	18	AAW22615	Murine 4 kb colony
24	123	5.7	552	21	AAE19547	Human colony stimu
25	123	5.7	552	21	AAE90329	Mouse CSF-1 protei
26	123	5.7	552	22	AAE00540	Murine colony stim
27	123	5.7	552	22	AAE59606	Murine colony stim
28	123	5.7	552	22	AAE49459	Murine colony stim
29	123	5.7	552	23	ABB57255	Murine CSF-1 #1.
30	121.5	5.6	2977	22	ABB69480	Mouse ischaemic co
31	118	5.4	377	17	AAW05411	Drosophila melanog
32	117	5.4	486	21	AAE43338	Human H74 protein.
33	117	5.4	486	21	AAE4041	Human ORFX ORF3102
34	115	5.3	783	23	AAE82709	Protein encoded by
35	115	5.3	1030	22	ABG23699	Amino acid sequenc
36	114	5.2	552	9	AAE80766	Novel human diagno
37	114	5.2	552	17	AAW10071	Deduced sequence f
38	114	5.2	552	18	AAW36146	Murine short form
39	114	5.2	552	18	AAW35753	Mouse short form o
40	114	5.2	552	18	AAW22616	Murine colony stim
41	114	5.2	552	21	AAE19548	Murine 2 kb colony
42	114	5.2	552	21	AAE90330	Human colony stimu
43	114	5.2	552	21	AAE90330	Mouse CSF-1 protei
44	114	5.2	552	22	AAE00541	Murine colony stim
45	114	5.2	552	22	AAE59607	Murine colony stim
			552	22	AAE49460	Murine CSF-1 #2.

ALIGNMENTS

RESULT 1
AAB03880
ID AAB03880 standard; protein; 414 AA.
XX
AC AAB03880;
XX
DT 26-OCT-2000 (first entry)
XX
DE Human ART-1 protein sequence.
XX
DE DE
KW Human: tumour antigen protein; ART-1; HLA antigen; cytostatic;
KW bound cytotoxic T cell; tumour; cancer; remedy; prevent; diagnose.
XX
OS Homo sapiens.
XX
PN WO2000032770-A1.
XX
PD 08-JUN-2000.
XX
PF 30-NOV-1999; 99WO-JP06682.
XX
PR 01-DEC-1998; 98JP-0341253.
XX
PA (SUMU) SUMITOMO PHARM CO LTD.
XX (ITOH/) ITOH K.
XX
PI Itoh K, Gomi S;
XX
DR WPI: 2000-412318/35.
XX N-PSDB; AAA62864.
XX
PT Novel tumor antigen protein ART-1, tumor antigen peptide originating from it, their derivatives, and DNAs, applicable in vivo or in vitro as

PT remedies, preventives and diagnostics for tumors -
XX Claim 1; Page 44-46; 59pp; Japanese.
XX The invention relates to a novel human tumour antigen protein, ART-1.
CC Included in the invention are polynucleotide sequences encoding the ART-1
CC protein, and mutated ART-1 proteins which when broken down
CC intracellularly produce a tumour antigen peptide that can recognise HLA
CC antigen and bound cytotoxic T cells. Antibodies which specifically
CC recognise ART-1 and its derivative peptides, are also included in the
CC invention. ART-1 exhibits cytostatic activity. The tumour antigen
CC protein, tumour antigen peptide originating from it, their derivatives,
CC and DNAs are applicable in vivo or in vitro as remedies, preventives and
CC diagnostics for tumors.
XX The present sequence represents the human ART-1 protein sequence.

XX Sequence 414 AA;
Query Match 100.0%; Score 2175; DB 21; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.4e-191;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNLQRYWGEIPISSQTNRRSFDLLPREFRLVEVDHPPHQPSPKPKPTMLDIPSEPC 60
DB 1 MNLQRYWGEIPISSQTNRRSFDLLPREFRLVEVDHPPHQPSPKPKPTMLDIPSEPC 60
QY 61 SLTHTIQLIOHNRRLNLIATAQAOQOQTEGVKTESEPLPSCPGSPPLPDDLPLDC 120
DB 61 SLTHTIQLIOHNRRLNLIATAQAOQOQTEGVKTESEPLPSCPGSPPLPDDLPLDC 120
QY 121 KNPAPFOIRHSDPESDFYRGKGPEVTELSWHSCRQLLYQAVATILAHAGFDCANESVLE 180
DB 121 KNPAPFOIRHSDPESDFYRGKGPEVTELSWHSCRQLLYQAVATILAHAGFDCANESVLE 180
QY 181 TLTDVAHEYCYLKFLLRFAVDREARLGQTFFPDVMEQVFEHVGIGSVLSQKFWQHRK 240
DB 181 TLTDVAHEYCYLKFLLRFAVDREARLGQTFFPDVMEQVFEHVGIGSVLSQKFWQHRK 240
QY 241 DYHSMYLIQSKOLSEEEYRIVNPEKATEDAPVKIKEPVSDFITFPVSEELADLASGDQ 300
DB 241 DYHSMYLIQSKOLSEEEYRIVNPEKATEDAPVKIKEPVSDFITFPVSEELADLASGDQ 300
QY 301 SLPMGVLGAQSERPPSNLEVEASQASSAEVNASPLNLAHVMEQVFEHVGIGSVLSQKFWQHRK 360
DB 301 SLPMGVLGAQSERPPSNLEVEASQASSAEVNASPLNLAHVMEQVFEHVGIGSVLSQKFWQHRK 360
QY 361 GSDVFEPMGMSAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQCKKMRKI 414
DB 361 GSDVFEPMGMSAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQCKKMRKI 414

RESULT 2
AAM25845
ID AAM25845 standard; Protein; 419 AA.
XX AAM25845;
XX 16-OCT-2001 (first entry)
XX Human protein sequence SEQ ID NO:1360.
XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiagregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antipressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;

KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.
XX Homo sapiens.
XX WO200153455-A2.
PD 26-JUL-2001.
XX 22-DEC-2000; 2000WO-US35017.
XX 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-457603/49.
DR N-PSDB; AAH99786.
XX Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
PS Claim 20; Page 281; 1217pp; English.
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnerary;
CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antipressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.

XX Sequence 419 AA;
Query Match 100.0%; Score 2175; DB 22; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.4e-191;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNLQRYWGEIPISSQTNRRSFDLLPREFRLVEVDHPPHQPSPKPKPTMLDIPSEPC 60
DB 6 MNLQRYWGEIPISSQTNRRSFDLLPREFRLVEVDHPPHQPSPKPKPTMLDIPSEPC 65
QY 61 SLTHTIQLIOHNRRLNLIATAQAOQOQTEGVKTESEPLPSCPGSPPLPDDLPLDC 120
DB 66 SLTHTIQLIOHNRRLNLIATAQAOQOQTEGVKTESEPLPSCPGSPPLPDDLPLDC 125
QY 121 KNPAPFOIRHSDPESDFYRGKGPEVTELSWHSCRQLLYQAVATILAHAGFDCANESVLE 180
DB 126 KNPAPFOIRHSDPESDFYRGKGPEVTELSWHSCRQLLYQAVATILAHAGFDCANESVLE 185
QY 181 TLTDVAHEYCYLKFLLRFAVDREARLGQTFFPDVMEQVFEHVGIGSVLSQKFWQHRK 240
DB 186 TLTDVAHEYCYLKFLLRFAVDREARLGQTFFPDVMEQVFEHVGIGSVLSQKFWQHRK 245
QY 241 DYHSMYLIQSKOLSEEEYRIVNPEKATEDAPVKIKEPVSDFITFPVSEELADLASGDQ 300

Db 246 DYHSYMLQISKQLSEYERIVNPEKATEAKPVYKKEEPPVSDITFFVSELEADLSDGQ 305
 QY 301 SLPWGLVGAQSERFPNLEVEASPOASSAEVNASPLWNLAHVKMEQSEEGNVSGHVL 360
 Db 306 SLPWGLVGAQSERFPNLEVEASPOASSAEVNASPLWNLAHVKMEQSEEGNVSGHVL 365
 QY 361 GSDVFEPPMGSEAGIPQSPDDSSYSGSHSDSLMGSPPVFNQCKKRMRI 414
 Db 366 GSDVFEPPMGSEAGIPQSPDDSSYSGSHSDSLMGSPPVFNQCKKRMRI 419
 RESULT 3
 AAG98644
 ID AAG98644 standard; Protein; 412 AA.
 AC AAG98644;
 XX
 XX
 DT 21-SEP-2001 (first entry)
 DE Human cell death protective cDNA clone CNI-00714 ORF1 protein, SEQ:77.
 XX
 XX Cell death protective; apoptosis; necrosis; human; drug screening;
 KW cell death-associated disorder; central nervous system disorder;
 KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
 KW stroke; cerebral infarction; ischaemic encephalopathy;
 KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
 KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
 KW vascular disease; ophthalmological disorder; diabetic retinopathy;
 KW macular degeneration; hypertension; myocardial infarction;
 KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
 KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
 KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
 KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
 KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
 KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
 KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.
 XX
 OS Homo sapiens.
 PN WO200145638-A2.
 XX
 XX 28-JUN-2001.
 PD
 XX 11-DEC-2000; 2000WO-US33547.
 PF
 XX 14-DEC-1999; 99US-0461697.
 PR
 XX (COGE-) COGENT NEUROSCIENCE INC.
 PA
 XX Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
 PI WPI: 2001-390297/41.
 XX N-PSDB; AAH84170, AAH84171.
 DR
 XX Novel protective sequence polynucleotides and polypeptides, used to
 PT identify modulators of their expression and activity, which are used in
 PT to treat central nervous system conditions, diseases and disorders -
 XX
 PS Claim 1; Fig 6A; 325pp; English.
 XX
 CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
 CC protect against cell death (i.e., apoptosis or necrosis). Sequences
 CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
 CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
 CC while the remaining nucleic acid sequences within the range given above
 CC represent the open reading frames (ORFs) of these cDNA clones. Sequences
 CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death
 CC protective ORFs. The cell death protective cDNA clones are able to
 CC prevent, delay or reverse progression through the apoptotic or necrotic
 CC pathways when injected into a cell predisposed to or undergoing cell
 CC death. The cell death protective nucleic acids and polypeptides can be
 CC used in the diagnosis and treatment of disorders associated with cell

CC death, and to screen for compounds which modulate their activity or
 CC expression. Such modulators, preferably a small organic molecule, an
 CC antibody, a ribozyme, or an antisense molecule, can also be used to treat
 CC cell death-related diseases. Such diseases include those associated with
 CC the central nervous system including psychiatric or neurological
 CC disorders, especially ischaemia-related conditions such as strokes, and
 CC also includes neurodegenerative disorders such as Alzheimer's disease,
 CC Huntington's disease, or Parkinson's disease. The modulators may also be
 CC used to treat infections such as meningitis, malaria, or trypanosomiasis;
 CC vascular diseases such as ischaemic encephalopathy or cerebral
 CC infarction; eye conditions such as diabetic retinopathy or macular
 CC degeneration; hypertension; myocardial infarction; atherosclerosis;
 CC respiratory conditions such as asthma or chronic obstructive pulmonary
 CC disease; neoplastic conditions such as cancers or benign tumours; blood
 CC cell conditions such as anaemia; gastrointestinal conditions such as
 CC gastritis or ulcerative colitis; liver conditions such as biliary
 CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;
 CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's
 CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune
 CC system disorders such as acquired immunodeficiency syndrome (AIDS). The
 CC nucleic acids may additionally be used to generate animal models of
 CC cell death-associated disorders. The present sequence represents a
 CC cell death protective polypeptide.
 XX
 SQ Sequence 412 AA;
 Query Match 99.1%; Score 2155; DB 22; Length 412;
 Best Local Similarity 100.0%; Pred. No. 9.8e-190;
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 RYWGEPISSTQNRSSFDLLPREFLVEVHPDPLHQPSANKPKPTMLDIPSECSLTI 64
 Db 3 RYWGEPISSTQNRSSFDLLPREFLVEVHPDPLHQPSANKPKPTMLDIPSECSLTI 62
 QY 65 HTIQLIHNRLRLNIATAQAQNOQTEGVKTESEPLSCPGSPPLDDLLPCKKNPN 124
 Db 63 HTIQLIHNRLRLNIATAQAQNOQTEGVKTESEPLSCPGSPPLDDLLPCKKNPN 122
 QY 125 APFQIRHSDPESDFYRGKGPVTELSWHSRCROLLQAVATILAHAGFDCANESVLETLTD 184
 Db 123 APFQIRHSDPESDFYRGKGPVTELSWHSRCROLLQAVATILAHAGFDCANESVLETLTD 182
 QY 185 VAHEYCLKFTKLLRFADVREARLGOTPPFDVMEQVFHEVGIGSVLSLQKFWQHRDKYHS 244
 Db 183 VAHEYCLKFTKLLRFADVREARLGOTPPFDVMEQVFHEVGIGSVLSLQKFWQHRDKYHS 242
 QY 245 YMLQISKQLSEYERIVNPEKATEAKPVYKKEEPPVSDITFFVSELEADLSDGQSLPM 304
 Db 243 YMLQISKQLSEYERIVNPEKATEAKPVYKKEEPPVSDITFFVSELEADLSDGQSLPM 302
 QY 305 GVLGAQSERFPNLEVEASPOASSAEVNASPLWNLAHVKMEQSEEGNVSGHVLGSDV 364
 Db 303 GVLGAQSERFPNLEVEASPOASSAEVNASPLWNLAHVKMEQSEEGNVSGHVLGSDV 362
 QY 365 FEPPMGSEAGIPQSPDDSSYSGSHSDSLMGSPPVFNQCKKRMRI 414
 Db 363 FEPPMGSEAGIPQSPDDSSYSGSHSDSLMGSPPVFNQCKKRMRI 412
 RESULT 4
 AAB45216
 ID AAB45216 standard; Protein; 403 AA.
 XX
 AC AAB45216;
 XX
 DT 12-FEB-2001 (first entry)
 XX
 DE Gene 41 human secreted protein homologous amino acid sequence #157.
 XX
 XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
 KW neurotropic; neuroprotective; antibacterial; virucide; fungicide; cancer;
 KW ophthalmological; autoimmune disease; hyperproliferative disorder;

KW cardiovascular disorder; cerebrovascular disorder; wound healing;
KW nervous system disorder; aging; chemotaxis.
OS Homo sapiens.
XX WO200058467-A1.
XX 05-OCT-2000.
XX 22-MAR-2000; 2000WO-US07505.
XX 26-MAR-1999; 99US-0126502.
XX 17-DEC-1999; 99US-0172410.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI; 2000-611712/58.
XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX Disclosure; Page 60-61; 440pp; English.
XX Polynucleotide sequences AAC80531-C80580 represent cDNA encoding human
CC secreted proteins AAB45120-B45169. Sequences AAB45170-B45225 represent
CC alternative polypeptides encoded by the genes, and amino acid sequences
CC to which they are homologous. The genes and proteins have activities
CC dependent on the tissues and cells in which they are expressed. Examples
CC of their activities include immunosuppressive; antiarthritic;
CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
CC cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
CC fungicide; and ophthalmological. The secreted proteins, polynucleotides,
CC antagonists and agonists may be useful in treating, preventing and/or
CC diagnosing diseases and disorders such as autoimmune diseases
CC e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms,
CC of the breast or liver, cerebrovascular disorders e.g. cardiac arrest,
CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous
CC system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC the invention.
XX used in the isolation and characterisation of the genes and proteins of
XX the invention.
XX Sequence 403 AA;
Query Match 97.5%; Score 2121; DB 21; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.3e-186;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 RWGPIPTSSQTNRSSFDLLPREFLVEHDPPLHQPANKPKPPTMLDIPSPCSLTI 64
DB 1 RWGPIPTSSQTNRSSFDLLPREFLVEHDPPLHQPANKPKPPTMLDIPSPCSLTI 60
QY 65 HTIQLIQRNRLNLATAQANQQQTGKTESEPLPCGSPPLDILLDCNPN 124
DB 61 HTIQLIQRNRLNLATAQANQQQTGKTESEPLPCGSPPLDILLDCNPN 120
QY 125 APFQIRHSDPSDFYRGKEPVTLSWHSCRLQLYQAVATTLAHAGFCANESVLEITD 184
DB 121 APFQIRHSDPSDFYRGKEPVTLSWHSCRLQLYQAVATTLAHAGFCANESVLEITD 180
QY 185 VAHEYCLAKFTLLRFVADREARLGQTPPDVMEQVFEHVGISVLSLQKFWQRIKDYHS 244
DB 181 VAHEYCLAKFTLLRFVADREARLGQTPPDVMEQVFEHVGISVLSLQKFWQRIKDYHS 240

QY 245 YMLQISKLSSEYERIVNPEKATEDAKPVKIKEEPVSDITPPVSEELASDQSLPM 304
DB YMLQISKLSSEYERIVNPEKATEDAKPVKIKEEPVSDITPPVSEELASDQSLPM 300
QY 305 GVLGAQSERFSPNLEVEASPOASSAEVNASPLWNLAHVKMEPQSEEGNVSGHVLGSDV 364
DB GVLGAQSERFSPNLEVEASPOASSAEVNASPLWNLAHVKMEPQSEEGNVSGHVLGSDV 360
QY 365 FEPMGSMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNORC 407
DB FEPMGSMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNORC 403
RESULT 5
AAB45217
ID AAB45217 standard; Protein; 404 AA.
XX
AC AAB45217;
XX
DT 12-FEB-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 41 SEQ ID NO:158.
XX
KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; autoimmune disease; hyperproliferative disorder;
KW cerebrovascular disorder; cerebrovascular disorder; wound healing;
KW nervous system disorder; aging; chemotaxis.
OS Homo sapiens.
PN WO200058467-A1.
XX
PD 05-OCT-2000.
XX
PF 22-MAR-2000; 2000WO-US07505.
XX
PR 26-MAR-1999; 99US-0126502.
XX
PR 17-DEC-1999; 99US-0172410.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI; 2000-611712/58.
XX N-PSDB; AAC80571.
XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX Disclosure; Page 60-61; 440pp; English.
XX Polynucleotide sequences AAC80531-C80580 represent cDNA encoding human
CC secreted proteins AAB45120-B45169. Sequences AAB45170-B45225 represent
CC alternative polypeptides encoded by the genes, and amino acid sequences
CC to which they are homologous. The genes and proteins have activities
CC dependent on the tissues and cells in which they are expressed. Examples
CC of their activities include immunosuppressive; antiarthritic;
CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
CC cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
CC fungicide; and ophthalmological. The secreted proteins, polynucleotides,
CC antagonists and agonists may be useful in treating, preventing and/or
CC diagnosing diseases and disorders such as autoimmune diseases
CC e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms,
CC of the breast or liver, cerebrovascular disorders e.g. cardiac arrest,
CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous
CC system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities. AAC80522-C80530 and AAB45119 represent sequences
CC used in the isolation and characterisation of the genes and proteins of
XX the invention.
XX Sequence 403 AA;

Db 181 QKFWHRTKDYHVMLOLQSKLSEYERIVNPEKATEDAKPVKKEEVPVSDITFPVSEEL 240
 QY 292 EADLASGQSLPVMGVLGAQSERFNSNLEVEASPOASSAEVNASPLWNLHVKMPEQSEEE 351
 Db 241 EADLASGQSLPVMGVLGAQSERFNSNLEVEASPOASSAEVNASPLWNLHVKMPEQSEEE 300
 QY 352 GNVSGHGVGLSDVFEPMGMSERAGIPQSPDSDSSYSGSHSTDSLMGSSPVFNQCKRKM 411
 Db 301 GNVSGHGVGLSDVFEPMGMSERAGIPQSPDSDSSYSGSHSTDSLMGSSPVFNQCKRKM 360
 QY 412 RKI 414
 Db 361 RKI 363

RESULT 7

AG98652
 ID AG98652 standard; Protein; 199 AA.

AC AG98652;

DT 21-SEP-2001 (first entry)

DE Human cell death protective cDNA clone CNI-00714 ORF9 protein, SEQ:93.

XX Cell death protective; apoptosis; necrosis; human; drug screening;
 KW cell death-associated disorder; central nervous system disorder;
 KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
 KW stroke; cerebral infarction; ischaemic encephalopathy;
 KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
 KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
 KW vascular disease; ophthalmological disorder; diabetic retinopathy;
 KW macular degeneration; hypertension; myocardial infarction;
 KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
 KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
 KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
 KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
 KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
 KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
 KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.

OS Homo sapiens.

XX WO200145638-A2.

PN 28-JUN-2001.

PD 11-DEC-2000; 2000WO-US33547.

PF 14-DEC-1999; 99US-0461697.

PR (COGE-) COGENT NEUROSCIENCE INC.

PA Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;

PI WPI; 2001-390297/41.

DR N-PSDB: AAH84170, AAH84179.

XX Novel protective sequence polynucleotides and polypeptides, used to
 PT identify modulators of their expression and activity, which are used in
 PT to treat central nervous system conditions, diseases and disorders -

PS Claim 1; Fig 6I; 325pp; English.

XX Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
 CC protect against cell death (i.e., apoptosis or necrosis). Sequences
 CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84226, AAH84265,
 CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
 CC while the remaining nucleic acid sequences within the range given above
 CC represent the open reading frames (ORFs) of these cDNA clones. Sequences
 CC AG98610-AG98829 represent the polypeptides encoded by the cell death
 CC protective ORFs. The cell death protective cDNA clones are able to
 CC prevent, delay or reverse progression through the apoptotic or necrotic

CC pathways when injected into a cell predisposed to or undergoing cell
 CC death. The cell death protective nucleic acids and polypeptides can be
 CC used in the diagnosis and treatment of disorders associated with cell
 CC death, and to screen for compounds which modulate their activity or
 CC expression. Such modulators, preferably a small organic molecule, an
 CC antibody, a ribozyme, or an antisense molecule, can also be used to treat
 CC cell death-related diseases. Such diseases include those associated with
 CC the central nervous system including psychiatric or neurological
 CC disorders, especially ischaemia-related conditions such as strokes, and
 CC also includes neurodegenerative disorders such as Alzheimer's disease,
 CC Huntington's disease, or Parkinson's disease. The modulators may also be
 CC used to treat infections such as meningitis, malaria, or trypanosomiasis;
 CC vascular diseases such as ischaemic encephalopathy or cerebral
 CC infarction; eye conditions such as diabetic retinopathy or macular
 CC degeneration; hypertension; myocardial infarction; atherosclerosis;
 CC respiratory conditions such as asthma or chronic obstructive pulmonary
 CC disease; neoplastic conditions such as cancers or benign tumours; blood
 CC cell conditions such as anaemia; gastrointestinal conditions such as
 CC gastritis or ulcerative colitis; liver conditions such as biliary
 CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;
 CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's
 CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune
 CC system disorders such as acquired immunodeficiency syndrome (AIDS). The
 CC nucleic acids may additionally be used to generate animal models of
 CC cell death-associated disorders. The present sequence represents a
 CC cell death protective polypeptide.

XX Sequence 199 AA;

Query Match 47.0%; Score 1022; DB 22; Length 199;
 Best Local Similarity 100.0%; Pred. No. 8.4e-86;
 Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 MEQVFHEVGVGSLVLSLQKFWQHRKTDYHVMLOLQSKLSEYERIVNPEKATEDAKPVKI 275
 Db 1 MEQVFHEVGVGSLVLSLQKFWQHRKTDYHVMLOLQSKLSEYERIVNPEKATEDAKPVKI 60
 QY 276 KEEPVSDITFPVSEELADLASGQSLPVMGVLGAQSERFNSNLEVEASPOASSAEVNASP 335
 Db 61 KEEPVSDITFPVSEELADLASGQSLPVMGVLGAQSERFNSNLEVEASPOASSAEVNASP 120
 QY 336 LWNLAHVKMEPQSEEGNVSGHGVGLSDVFEPMGMSERAGIPQSPDSDSSYSGSHSTDS 395
 Db 121 LWNLAHVKMEPQSEEGNVSGHGVGLSDVFEPMGMSERAGIPQSPDSDSSYSGSHSTDS 180
 QY 396 LMGSSPVFNQCKRMRKI 414
 Db 181 LMGSSPVFNQCKRMRKI 199

RESULT 8

AA55941
 ID AA55941 standard; Protein; 198 AA.

XX AA55941;

AC 31-JAN-2000 (first entry)

DT Human endometrium tumour EST encoded protein 1.

DE Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;
 KW treatment; uterine; gene therapy; expressed sequence tag.

OS Homo sapiens.

XX DEL9817948-A1.

XX 21-OCT-1999.

PF 17-APR-1998; 98DE-1017948.

XX 17-APR-1998; 98DE-1017948.

XX 17-APR-1998; 98DE-1017948.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
PI WPI: 1999-591957/51.
XX N-PSDB; AAZ41981.
XX
XX New nucleic acid sequences expressed in uterine cancer tissues, and
XX derived polypeptides, for treatment of uterine and endometrial cancer
XX and identification of therapeutic agents -
XX
XX Claim 23; Page 275; 444pp; German.
XX
XX This invention describes novel human nucleic acid (cDNA) sequences (A),
XX that are highly expressed in uterine tumour tissue and which have
XX anticancer and cytostatic activity. (A) are used (i) for recombinant
XX expression of polypeptides (B) and (ii) to isolate complete genes. (B)
XX are used (i) to identify agents suitable for treatment of uterine or
XX endometrial cancer; (ii) directly for treating these forms of cancer
XX (including expression from gene therapy vectors) and (iii) for
XX generation of specific antibodies. (A) are identified by assembling ESTs
XX (expressed sequence tags) from a particular tissue type before comparison
XX of expression patterns. This allows a significantly longer fragment of
XX the gene to be revealed, so should reduce the number of failures
XX associated with the fact that ESTs from different libraries may represent
XX different parts of the same unknown gene, distorting the estimated
XX frequency of occurrence in a particular tissue. AAY59941-Y60328 represent
XX protein fragments encoded by the human endometrium tumour cDNA library
XX derived EST fragments represented in AAZ41981-242121.
XX
XX Sequence 198 AA;
XX
XX Query Match 42.1%; Score 915; DB 20; Length 198;
XX Best Local Similarity 91.6%; Pred. No. 6.1e-76;
XX Matches 175; Conservative 1; Mismatches 13; Indels 2; Gaps 2;
QY 1 MNLQRYWGEIPISSSQTNRSFLLPREFLVEVHDPLHOPSAKPKPTMLDIPSEPC 60
DB 6 MNLQRYWGEIPISSSQTNRSFLLPREFLVEVHDPLHOPSAKPKPTMLDIPSEPC 65
QY 61 SLTIHTLIQIHRNRLNLTAAQAQNOQOQTEGVKTEESPLPSCGSPPLPDDLPLDC 120
DB 66 SLTIHTLIQIHRNRLNLTAAQAQNOQOQTEGVKTEESPLPSCGSPPLPDDLPLDC 125
QY 121 KPNAPFOIRHSPEDSFYRGKPEVTELSWHSCROLLYQAVATILA-HAGFCANESVL 179
DB 126 KPNAPFOIRHSPEDSFYRGKPEVTELSWHSCROLLYOGSGTNPQORRAFCANESVL 185
QY 180 ETLT-DVAHEY 189
DB 186 EDPNMLAHEY 196
RESULT 9
AAG98634
ID AAG98654 standard; Protein: 169 AA.
XX
XX AAG98654;
XX
XX 21-SEP-2001 (first entry)
XX
XX Human cell death protective cDNA clone CNI-00714 ORF11 protein, SRQ:97.
DE
XX Cell death protective; apoptosis; necrosis; human; drug screening;
XX cell death-associated disorder; central nervous system disorder;
XX psychiatric disorder; neurological disorder; ischaemia-related disorder;
XX stroke; cerebral infarction; ischaemic encephalopathy;
XX neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
XX Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
XX vascular disease; ophthalmological disorder; diabetic retinopathy;
XX macular degeneration; hypertension; myocardial infarction;
XX atherosclerosis; respiratory disorder; asthma; transgenic animal;
XX chronic obstructive pulmonary disease; neoplastic condition; cancer;

KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.
XX
OS Homo sapiens.
PN WO200145638-A2.
XX
PD 28-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33547.
XX
PR 14-DEC-1999; 99US-0461697.
XX
PA (COGE-) COGENT NEUROSCIENCE INC.
XX
PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
XX WPI: 2001-390297/41.
DR N-PSDB; AAH84170, AAH84181.
XX
XX Novel protective sequence polynucleotides and polypeptides, used to
XX identify modulators of their expression and activity, which are used in
XX to treat central nervous system conditions, diseases and disorders -
XX
XX Claim 1; Fig 6K; 325pp; English.
XX
XX Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
XX protect against cell death (i.e., apoptosis or necrosis). Sequences
XX AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
XX AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
XX while the remaining nucleic acid sequences within the range given above
XX represent the open reading frames (ORFs) of these cDNA clones. Sequences
XX AAG98610-AAG98829 represent the polypeptides encoded by the cell death
XX protective ORFs. The cell death protective cDNA clones are able to
XX prevent, delay or reverse progression through the apoptotic or necrotic
XX pathways when injected into a cell predisposed to or undergoing cell
XX death. The cell death protective nucleic acids and polypeptides can be
XX used in the diagnosis and treatment of disorders associated with cell
XX death, and to screen for compounds which modulate their activity or
XX expression. Such modulators, preferably a small organic molecule, an
XX antibody, a ribozyme, or an antisense molecule, can also be used to treat
XX cell death-related diseases. Such diseases include those associated with
XX the central nervous system including psychiatric or neurological
XX disorders, especially ischaemia-related conditions such as strokes, and
XX also includes neurodegenerative disorders such as Alzheimer's disease,
XX Huntington's disease, or Parkinson's disease. The modulators may also be
XX used to treat infections such as meningitis, malaria, or trypanosomiasis;
XX vascular diseases such as ischaemic encephalopathy or cerebral
XX infarction; eye conditions such as diabetic retinopathy or macular
XX degeneration; hypertension; myocardial infarction; atherosclerosis;
XX respiratory conditions such as asthma or chronic obstructive pulmonary
XX disease; neoplastic conditions such as cancers or benign tumours; blood
XX cell conditions such as anaemia; gastrointestinal conditions such as
XX gastritis or ulcerative colitis; liver conditions such as biliary
XX cirrhosis; kidney disorders such as glomerulonephritis; cystitis;
XX endometriosis; endocrine disorders such as Grave's disease or Hashimoto's
XX thyroiditis; skin conditions such as dermatitis or urticaria; or immune
XX system disorders such as acquired immunodeficiency syndrome (AIDS). The
XX nucleic acids may additionally be used to generate animal models of
XX cell death-associated disorders. The present sequence represents a
XX cell death protective polypeptide.
XX
XX Sequence 169 AA;
XX
XX Query Match 39.4%; Score 858; DB 22; Length 169;
XX Best Local Similarity 100.0%; Pred. No. 8.5e-71;
XX Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 MLOISKOLSEYERIVNPEKATEDAKPVKIEEVPVSDITFPVSEELADIASGDSLPMG 305
|||||

XX PA (COGE-) COGENT NEUROSCIENCE INC.
XX KW Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
XX KW WPI; 2001-390297/41.
XX DR N-PSDB; AAH84170, AAH84185.
XX XX
XX PT Novel protective sequence polynucleotides and polypeptides, used to
XX PT identify modulators of their expression and activity, which are used in
XX PT to treat central nervous system conditions, diseases and disorders -
XX PS Claim 1; Fig 60; 325pp; English.
XX XX
XX CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
XX CC protect against cell death (i.e., apoptosis or necrosis). Sequences
XX CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
XX CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
XX CC while the remaining nucleic acid sequences within the range given above
XX CC represent the open reading frames (ORFs) of these cDNA clones. Sequences
XX CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death
XX CC protective ORFs. The cell death protective cDNA clones are able to
XX CC prevent, delay or reverse progression through the apoptotic or necrotic
XX CC pathways when injected into a cell predisposed to or undergoing cell
XX CC death. The cell death protective nucleic acids and polypeptides can be
XX CC used in the diagnosis and treatment of disorders associated with cell
XX CC death, and to screen for compounds which modulate their activity or
XX CC expression. Such modulators, preferably a small organic molecule, an
XX CC antibody, a ribozyme, or an antisense molecule, can also be used to treat
XX CC cell death-related diseases. Such diseases include those associated with
XX CC the central nervous system including psychiatric or neurological
XX CC disorders, especially ischaemia-related conditions such as strokes, and
XX CC also includes neurodegenerative disorders such as Alzheimer's disease,
XX CC Huntington's disease, or Parkinson's disease. The modulators may also be
XX CC used to treat infections such as meningitis, malaria, or trypanosomiasis;
XX CC vascular diseases such as ischaemic encephalopathy or cerebral
XX CC infarction; eye conditions such as diabetic retinopathy or macular
XX CC degeneration; hypertension; myocardial infarction; atherosclerosis;
XX CC respiratory conditions such as asthma or chronic obstructive pulmonary
XX CC disease; neoplastic conditions such as cancers or benign tumours; blood
XX CC cell conditions such as anaemia; gastrointestinal conditions such as
XX CC gastritis or ulcerative colitis; liver conditions such as biliary
XX CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;
XX CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's
XX CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune
XX CC system disorders such as acquired immunodeficiency syndrome (AIDS). The
XX CC nucleic acids may additionally be used to generate animal models of
XX CC cell death-associated disorders. The present sequence represents a
XX CC cell death protective polypeptide.
XX SQ Sequence 71 AA;
Query Match 17.1%; Score 371; DB 22; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.7e-26;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 344 MEPQSEEGNVSHGVLSDFEPMGMSGAEIGPQSDSSYGSHTDSLGMSSPVF 403
Db 1 MEPQSEEGNVSHGVLSDFEPMGMSGAEIGPQSDSSYGSHTDSLGMSSPVF 60
QY 404 NORCKRMKRI 414
Db 61 NORCKRMKRI 71
RESULT 12
AAG9862
ID AAG9862 standard; Protein; 46 AA.
XX AC AAG9862;
XX XX
XX DT 21-SEP-2001 (first entry)
XX XX

DE XX Human cell death protective cDNA clone CNI-00714 ORF19 protein, SEQ:113.
XX KW Cell death protective; apoptosis; necrosis; human; drug screening;
XX KW cell death-associated disorder; central nervous system disorder;
XX KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
XX KW stroke; cerebral infarction; ischaemic encephalopathy;
XX KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
XX KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
XX KW vascular degeneration; ophthalmological disorder; diabetic retinopathy;
XX KW macular degeneration; hypertension; myocardial infarction;
XX KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
XX KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
XX KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
XX KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
XX KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
XX KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
XX KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.
OS Homo sapiens.
XX PN WO200145638-A2.
XX XX
XX PD 28-JUN-2001.
XX XX
XX PF 11-DEC-2000; 2000WO-US33547.
XX XX
XX PR 14-DEC-1999; 99US-0461697.
XX XX
XX PA (COGE-) COGENT NEUROSCIENCE INC.
XX XX
XX PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
XX XX WPI; 2001-390297/41.
XX XX N-PSDB; AAH84170, AAH84189.
XX DR
XX XX
XX PT Novel protective sequence polynucleotides and polypeptides, used to
XX PT identify modulators of their expression and activity, which are used in
XX PT to treat central nervous system conditions, diseases and disorders -
XX PS Claim 1; Fig 65; 325pp; English.
XX XX
XX CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
XX CC protect against cell death (i.e., apoptosis or necrosis). Sequences
XX CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
XX CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
XX CC while the remaining nucleic acid sequences within the range given above
XX CC represent the open reading frames (ORFs) of these cDNA clones. Sequences
XX CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death
XX CC protective ORFs. The cell death protective cDNA clones are able to
XX CC prevent, delay or reverse progression through the apoptotic or necrotic
XX CC pathways when injected into a cell predisposed to or undergoing cell
XX CC death. The cell death protective nucleic acids and polypeptides can be
XX CC used in the diagnosis and treatment of disorders associated with cell
XX CC death, and to screen for compounds which modulate their activity or
XX CC expression. Such modulators, preferably a small organic molecule, an
XX CC antibody, a ribozyme, or an antisense molecule, can also be used to treat
XX CC cell death-related diseases. Such diseases include those associated with
XX CC the central nervous system including psychiatric or neurological
XX CC disorders, especially ischaemia-related conditions such as strokes, and
XX CC also includes neurodegenerative disorders such as Alzheimer's disease,
XX CC Huntington's disease, or Parkinson's disease. The modulators may also be
XX CC used to treat infections such as meningitis, malaria, or trypanosomiasis;
XX CC vascular diseases such as ischaemic encephalopathy or cerebral
XX CC infarction; eye conditions such as diabetic retinopathy or macular
XX CC degeneration; hypertension; myocardial infarction; atherosclerosis;
XX CC respiratory conditions such as asthma or chronic obstructive pulmonary
XX CC disease; neoplastic conditions such as cancers or benign tumours; blood
XX CC cell conditions such as anaemia; gastrointestinal conditions such as
XX CC gastritis or ulcerative colitis; liver conditions such as biliary
XX CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;
XX CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's
XX CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune
XX CC system disorders such as acquired immunodeficiency syndrome (AIDS). The
XX CC nucleic acids may additionally be used to generate animal models of
XX CC cell death-associated disorders. The present sequence represents a
XX CC cell death protective polypeptide.

CC nucleic acids may additionally be used to generate animal models of
CC cell death-associated disorders. The present sequence represents a
CC cell death protective polypeptide.
XX
SQ Sequence 46 AA;

Query Match 11.0%; Score 239; DB 22; Length 46;
Best Local Similarity 100.0%; Pred. No. 1.3e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 MSGMSEAGIPOSDDSSSYGSHSTDSLGMSSPVFNQCKRMKRI 414
|||||
DB 1 MSGMSEAGIPOSDDSSSYGSHSTDSLGMSSPVFNQCKRMKRI 46

RESULT 13
AAG98663
ID AAG98663 standard; Protein; 43 AA.
AC AAG98663;
DT
DT
DT
XX 21-SEP-2001 (first entry)
DE Human cell death protective cDNA clone CNI-00714 ORF20 protein, SEQ:115.
XX
KW Cell death protective; apoptosis; necrosis; human; drug screening;
KW cell death-associated disorder; central nervous system disorder;
KW psychiatric disorder; neurological disorder; ischaemia-related disorder;
KW stroke; cerebral infarction; ischaemic encephalopathy;
KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW vascular disease; ophthalmological disorder; diabetic retinopathy;
KW macular degeneration; hypertension; myocardial infarction;
KW atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.
XX
OS Homo sapiens.
XX
PN WO200145638-A2.
XX
PD
XX
XX 28-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33547.
XX
PR 14-DEC-1999; 99US-0461697.
XX
PA (COGE-) COGENT NEUROSCIENCE INC.
XX
PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;
XX
XX WPI; 2001-390297/41.
DR N-PSDB; AAH84170, AAH84190.
XX
XX Novel protective sequence polynucleotides and polypeptides, used to
PT identify modulators of their expression and activity, which are used in
PT to treat central nervous system conditions, diseases and disorders -
XX
PS Claim 1; Fig 6T; 325pp; English.
XX
XX Sequences AAH84132-AAH84370 represent human nucleic acid sequences which
CC protect against cell death (i.e., apoptosis or necrosis). Sequences
CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,
CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,
CC while the remaining nucleic acid sequences within the range given above
CC represent the open reading frames (ORFs) of these cDNA clones. Sequences
CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death
CC protective ORFs. The cell death protective cDNA clones are able to
CC prevent, delay or reverse progression through the apoptotic or necrotic

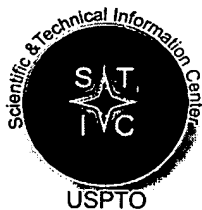
CC pathways when injected into a cell predisposed to or undergoing cell
CC death. The cell death protective nucleic acids and polypeptides can be
CC used in the diagnosis and treatment of disorders associated with cell
CC death, and to screen for compounds which modulate their activity or
CC expression. Such modulators, preferably a small organic molecule, an
CC antibody, a ribozyme, or an antisense molecule, can also be used to treat
CC cell death-related diseases. Such diseases include those associated with
CC the central nervous system including psychiatric or neurological
CC disorders, especially ischaemia-related conditions such as strokes, and
CC also includes neurodegenerative disorders such as Alzheimer's disease,
CC Huntington's disease, or Parkinson's disease. The modulators may also be
CC used to treat infections such as meningitis, malaria, or trypanosomiasis;
CC vascular diseases such as ischaemic encephalopathy or cerebral
CC infarction; eye conditions such as diabetic retinopathy or macular
CC degeneration; hypertension; myocardial infarction; atherosclerosis;
CC respiratory conditions such as asthma or chronic obstructive pulmonary
CC disease; neoplastic conditions such as cancers or benign tumours; blood
CC cell conditions such as anaemia; gastrointestinal conditions such as
CC gastritis or ulcerative colitis; liver conditions such as biliary
CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;
CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's
CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune
CC system disorders such as acquired immunodeficiency syndrome (AIDS). The
CC nucleic acids may additionally be used to generate animal models of
CC cell death-associated disorders. The present sequence represents a
XX cell death protective polypeptide.
SQ Sequence 43 AA;

Query Match 10.3%; Score 224; DB 22; Length 43;
Best Local Similarity 100.0%; Pred. No. 2.8e-13;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 MSEAGIPOSDDSSSYGSHSTDSLGMSSPVFNQCKRMKRI 414
|||||
DB 1 MSEAGIPOSDDSSSYGSHSTDSLGMSSPVFNQCKRMKRI 43

RESULT 14
ABB62430
ID ABB62430 standard; Protein; 359 AA.
XX
AC ABB62430;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 14082.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR N-PSDB; ABL06533.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX

THIS PAGE BLANK (USPTO)



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

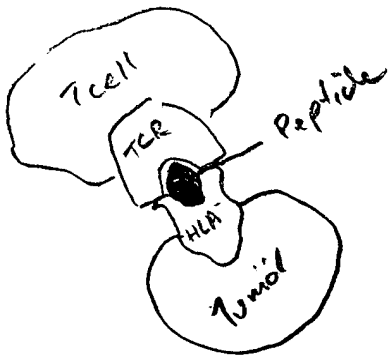
➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed form to STIC/Biotech-Chem Library CM1 – Circ. Desk





Tumor Ag From Squamous cell carcinoma

HLA-A24 or HLA-A26